Year : 2017

EVALUTATION OF SCIENTIFIC EVIDENCE : A PROPOSAL ON ONTOLOGICAL AND EPISTEMOLOGICAL BASES, AND SOME STATISTICAL APPLICATIONS

Lucena Molina José Juan

UNIL | Université de Lausanne
Faculté de droit,
des sciences criminelles
et d'administration publique

EVALUATION OF SCIENTIFIC EVIDENCE:

A PROPOSAL ON ONTOLOGICAL AND EPISTEMOLOGICAL BASES, AND SOME STATISTICAL APPLICATIONS

PhD THESIS of **José Juan Lucena Molina** established under the supervision of **Professor Franco Taroni**

Lausanne, 2017

ISBN

UNIL 1 Université de Lausanne
Ecole des sciences criminelles
batiment Batochime
CH-1015 Lausanne

IMPRIMATUR

A l'issue de la soutenance de these, le Jury autorise l'impression de la these de M. le Colonel José-Juan Lucena-Molina, candidat au doctorat en science forensique, intitulée

« Evaluation of scientific evidence : a proposal on ontological and epistemological bases, and sorne statistical applications »

Le Président du Jury

Professeur Marcelo Aebi

Lausanne, le 12 mai 2017

A mi padre, Coronel de la Guardia Civil, que falleció durante la redacción de esta tesis.

ACKNOWLEDGMENTS

I worked in the Criminalistics Service of the Spanish Civil Guard for 26 years and 4 months straight. This presence in the same unit is rather uncommon among the members of this military institution. I thank God for such a professional opportunity and all his ordinary Providence, Saint Josemaría Escrivá de Balaguer for teaching me that any honest job is a way of sanctification, and all the heads of the central laboratory responsible for the decision of my continuity each time I was promoted to a higher rank.

I am indebted to many scholars such as Franco Taroni, Alex Biedermann, Javier Ortega-García, Joaquín González-Rodríguez, Daniel Ramos-Castro, and others, because they helped me to write this doctoral thesis. They have been the main sources from which I have learnt about Forensic Science throughout my life.

I am very grateful to the Institute of the Scientific Police of the University of Lausanne for allowing me to submit this doctoral thesis.

I am especially grateful for those Civil Guards under my command which have been magnificent co-workers.

This multidisciplinary thesis would not be possible without the priceless assistance of some outstanding professors in philosophy such as Mariano Artigas and Juan José Sanguineti at the University of Navarra (Spain); David Teira, at the Spanish National Distance Education; Juan Arana at the University of Sevilla; and two professors of law: Marina Gascón at the University of Castilla-La Mancha and Virginia Pardo at the University of Valencia.

Lastly, I want to acknowledge my parents, six sisters and two brothers for their unconditional support carrying out this enterprise since the very beginning.

CONTENTS (pages 1-5)

PREFACE

ENFSI, the European Network of Forensic Science Institutes, was created in 1992 in an attempt to emulate ASCLD, the American Society of Crime Laboratory Directors, established in 1974. The first meeting organised by their promoters was held in Rijswijk, at the headquarters of the Netherlands Forensic Institute (NFI), in the spring of 1993. The foundational meeting was held in Wiesbaden, at the headquarters of the Bundeskriminalamt (BKA), in October 1995; and the first congress of the European Academy of Forensic Sciences (EAFS) was held at the University of Lausanne in September 1997.

ENFSI is currently made up of 67 laboratories from 36 countries. The main activity of ENFSI is carried out by their 17 working groups and the Standing Committees of Quality and Competence, and Research and Development. The Criminalistics Service of the Civil Guard joined ENFSI in 1998.

In 2009, ENFSI set in motion the first financial official announcement regarding research projects in forensic science by the European Commission. In the 2010 official announcement, the financing of the project titled “Development and implementation of an ENFSI standard for reporting evaluative forensic evidence” was approved. This project ended in December 2014 and achieved one of the most desired aims of the specialised forensic scientific community because the first technical guideline on evaluative conclusions in forensic reports within such a scope was published. Since the project was approved in 2010, every Annual Plan of the ENFSI Board has considered the project to be one of the most important aims of the organisation, with it being expressly mentioned as a specific aim year after year.

The current doctoral thesis is framed in this context. Its author belonged to the Criminalistics Service of the Civil Guard from 1988 to 2015. Ten years after the beginning of this period, in 1998, the second Odyssey conference was held in Avignon between 20 and 23 April, promoted by the then ESCA (European Speech Communication Association), and later ISCA when it became fully international. One of the speakers was Professor Champod and his presentation was titled “The inference of identity in Forensic Speaker Recognition”. The congress expressly mentioned its thematic opening to criminalistic applications, so that Professor Champod’s speech was made in a natural way. It was the first time the author of this thesis had personally heard about evaluation of scientific evidence from the Bayesian perspective.

The scientific collaboration between the Criminalistics Service of the Civil Guard and the ATVS (Speech and Signal Processing) research group, which at that time had been accountable to the Polytechnical University of Madrid since 1997, resulted in the first automatic speaker recognition system for forensic purposes in an official European laboratory, and was named IdentiVox. In 2004, AGNITIO S.L. was created as a spin-off company from the above-mentioned university, and the ATVS ASR technology was transferred to it. As the main novelty, the software enabled evaluation of the strength of the evidence in voice comparisons by means of a statistic called “likelihood ratio”. Up till then, ATVS, AGNITIO and the Criminalistics Service had held a strong scientific collaboration to validate the forensic procedure to make official reports. With this aim in mind, a number of databases were recorded for earlier years under different conditions of channel, sex, types of microphone, language and speech style. At the same time, ATVS took part in the NIST SRE (National Institute of Standards and Technology) evaluations, which had been sponsored by the US Department of Commerce since 2000.

Speaker variability due to intrinsic and extrinsic factors caused identity characteristics of voices to be modelled in very disadvantageous circumstances with respect to any other identity criminalistic evidence. From the very beginning, statistical techniques were necessary to compensate channels and normalise scores. On the other hand, an important research effort to create databases needed to be carried out due to the inter-speaker variability, as mentioned earlier, leading ATVS researchers to a scenario in which their main experimental effort for 2000–2003 was how to calculate robust likelihood ratios in forensic speaker recognition (Gonzalez et al., 2006, pp. 331–355).

The participation of the Criminalistics Service of the Civil Guard in the ENFSI Speech and Audio Analysis Working Group, and in consecutive meetings of the EAFS, allowed ATVS and Criminalistics Service members to present the advances experimented year after year using the developed tools for forensic speaker recognition. The commercial name given by AGNITIO to the IdentiVox system, BATVOX, started to be widely known in the European specialised scientific community, and the company began its expansion worldwide. Nowadays, on the official website of the company it is stated that the forensic product BATVOX is being used before the courts in more than 35 countries.

The acoustics area of the Criminalistics Service has gained a lot of experience defending expert reports on speaker recognition. Approximately 500 forensic reports in which likelihood ratios were assigned had been issued by the time this author left the area (2008). Though the Spanish judicial system allows official forensic experts to be cross-examined by private experts, in practice there have been few occasions when this has happened in this field. There have been many times that parties have reached an agreement; therefore, experts' appearances before the court have not been necessary. However, when appearances have taken place, experts have had the opportunity to explain what their expert activities have consisted of, thereby making the extent to which their conclusions are intelligible for those who heard them verifiable.

It may be said that courts reinterpret conclusions drawn by experts by means of likelihood ratios as if they were conclusions on probabilities of hypotheses alleged by parties. This is real life and it is a typical case of the well-known prosecutor's fallacy.

The above-mentioned illustrates the main purpose of this doctoral thesis, which is focused on the ontological and epistemological foundations of the proposal made by the specialised forensic community in favour of the use of Bayes' Theorem and Jeffrey's rule according to logic principles and laws to solve the problem of the evaluation of scientific evidence by courts.

The development of a specific technical guide on evaluative forensic conclusions has emphasised the need to use a glossary of terms acceptable both by two very different judicial systems in Europe, the Continental and the Anglo-Saxon ones, and by the special characteristics of the coexistence of so many languages in it. Terminological differences based on philosophical viewpoints were noted during the development of the project. These differences justify important semantic changes in words of current languages derived from the same Latin or Greek words, such as evidence, uncertainty, doubt, certainty, truth, probability or proof, which have an unquestionable importance when defending an expert report before the court.

From the judicial practice viewpoint, a peculiar circumstance is tackled in Spanish law when it considers scientific evidence such as the determination of the nature, weight and purity of

drugs in the brief penal proceedings as documentary evidence. This circumstance is interesting because the legal amendment was justified in considering such forensic reports to be indisputable data generators, with the forensic reports being assimilated to documents from the legal proceedings viewpoint. The extension of this point to other evidence types in some Spanish court sentences, and the possible future use of this procedural treatment in the ordinary legal proceedings due to some jurisprudential interpretation, caused the author of this thesis to produce a joint paper with other authors in 2010. This paper was finally published in the jurisprudence section of the *American Journal of Forensic Sciences* in July 2012.

Two researches linked to forensic practice have been carried out, the first one about validation of biometric methods when assigning likelihood ratios by means of using the so-called Limit Tippett plots, and the second one about the possible technical assistance for judges from forensic experts when assigning a priori odds applying Bayes' Theorem by means of Bayesian networks.

This thesis analyses several interdisciplinary aims: ontological, epistemological, judicial and logic-mathematical ones, all of them related to expert report conclusions.

It is believed that building a robust reasoning logic to make probabilistic inferences in forensic science from a merely mathematical, statistical or logistical viewpoint is not enough. Mathematical logic is the positive science of reasoning and as for that it is only interested in the positive calculus of its validity, regardless of any prior ontological assumption. But without a determined ontology and epistemology, which implies defining the concepts that they will use, it seems difficult that the proposed scientifically correct logical solution can be successful as a European standard because it has to be based on judicial language. Forensic experts and courts are not interested in the development of a positive science but in a practical science: in clarifying whether certain known facts are related to a possible crime. Therefore, not only is the coherence of the demonstrative logic reasoning used (logic of propositions and reasoning) important, but also the precision of the concepts used by language and consistency among them in reasoning (logic of concepts).

This author has taken every opportunity to speak about criminalistics at conferences for years, in general to underline that the conclusions of expert reports are that part of the expert's job that he/she can be sure will be unavoidably examined by the court's members. The need to make them precise in answering petitioners' questions implies a special difficulty. The relevance of the scientific evidence in the sentence indicates the importance of expressing conclusions according to the applied science and the rules of logic of reasoning under uncertainty even more.

If many forensic experts have emphasised the relevance of a closer scene of crime inspection they should spare no expense with the expert report conclusions. Undoubtedly, this last assertion is still a target beyond our reach in Europe, although the publication of a technical guide on evaluative conclusions lays the foundations for a paradigm shift in forensic science that some authors have described as the most important consequence of the advances made by that science in the 20th century (Evett, 2009, pp. 159–160).

Therefore, one of the aims of this doctoral thesis is to show the following original research points:

a) In the author's opinion, some especially relevant concepts used in the works of the main researches that defend the likelihood principle in forensic science are influenced by logical positivism. The current thesis studies some of these terms, used in those writings with a strongly biased semantic positivism, and compares their meanings with those used in the Aristotle-Thomas school. It is intended to show that for this philosophical school the subjective concept of probability is acceptable. However, it is pointed out that the confusion between the verbs 'to know' and 'to believe' that the author of this thesis perceives in the positivist justification theory implies an intrinsic difficulty in understanding and undoubtedly accepting the concept of subjective probability.

b) The validation of the pattern recognition methods applied to forensic science to calculate likelihood ratios from intra-variability and inter-variability distributions of normalised scores requires mechanisms to check the extent to which those values are well calibrated. A simple method, considered as being very useful for forensic experts, is offered to detect badly calibrated likelihood ratios assigned by a recognition or expert system that is proved with an experimental database. Basically it consists of using the classical Tippett plots, adding them the so-called Limit Tippett plots related to the probability of misleading evidence for each value of likelihood ratio obtained by the system.

c) It analyses the particular case of Spanish law in which scientific evidence on the determination of the nature, weight and purity of drugs has turned into documentary evidence in the brief penal proceedings. The Spanish legislator's justification for carrying out such a change is discussed and its arguments are refuted from a scientific viewpoint.

d) Finally, as a result of vast experience defending evaluative conclusions of forensic reports before the Spanish courts, a possible solution to the practical problem of courts or juries using the odds version of Bayes' Theorem is explored. The possible use of Bayesian networks to help courts and juries to assign a priori odds has been specifically explored.

Four different but complementary epistemological fields are harmonised in this doctoral thesis: ordinary knowledge, law, philosophy and science. They are indeed irreplaceable fields. It is believed that this interdisciplinary approach allows researchers to study the evaluative reporting in forensic science from an outstanding viewpoint.

PART I: THEORETICAL ASPECTS

Chapter 1. ONTOLOGICAL AND EPISTEMOLOGICAL BASES

1.1 Philosophy of being

Let me start with a brief introduction to the main principles of metaphysics and epistemology in the Aristotelian-Thomistic tradition. This will provide the foundations of the concept uncertainty I will use throughout the dissertation.

Aristotle defined metaphysics as 'first philosophy' with reference to its primary nature, because he considered that this part of philosophy was focused on the aspects of the real, which are the foundation of every single particular study. Traditionally, metaphysics is described as the study of being (from the Latin word *ens, entis*) as being, as well as the attributes of everything that is real (as distinct from what is imaginary or mental). Therefore, if metaphysics studies any real thing, that is, it is not focused on a specific being but on all kinds of realities, we may wonder: since uncertainty is a mental state, is it a material object of study for metaphysics? Insofar as uncertainty is a psychological effect of an imperfect act of knowledge, as will be explained further in detail, it has a certain being that is also studied by metaphysics.

Metaphysics claims that being is to intelligence what light is to the eyes. In order to understand something as real, being is needed. What is then uncertainty? As I have recently argued (Lucena, 2016, p. 150, fn. 28)[1]: "If we stick to the linguistic concept of uncertainty, we know it refers to a state of mind regarding the truth of a judgement. It deals with the intellectual operation performed by an individual capable of such a task and also with the results of the operation itself. Even more, it deals with the unsatisfactory result of such an operation as the intellect aspires to rest in the possession of truth when it recognises the reality."

Being involves not only the concepts but also the judgements we formulate. Our mind strengthens or rejects propositions, and by this it either adjusts to reality (then the proposition is true) or not (then the proposition is false). Every single judgement involves being as it eventually involves the very first of all judgements and the very simplest of them: being is not non-being. We say that our first judgement is the simplest because it is based on the most elemental concepts of the real: being and non-being. It is impossible to suppose the same thing to be and not to be at the same time and in the same sense, and this judgement was called the 'principle of non-contradiction'. This is the most self-evident proposition that can be expressed.

Following St. Thomas Aquinas' *Disputed questions on truth*, Forment states the genesis of the principle of non-contradiction as follows: "The notion of being is the first in the genesis of transcendental properties and first principles. The negation comes after this first affirmation and the concept of nothing appears. Third, an affirmation is made again and the concept of division is made up now, the separation between being and non-being. Its expression is the principle of non-contradiction: being is not non-being" (Forment, 2009, p. 276).

[1] Page numbers and footnotes of this reference will be written according to the pagination of this document because the paper has been included in Chapter 9.

The principle of non-contradiction is nearly always implicitly or indirectly used to reject what is incoherent, without being an explicit premise of each argument each time. Any proper knowledge has to be reached according to this principle (it's not the same to say 'according to' as 'from', Alvira et al., 2001, p. 48). This principle is the first premise of the truth of any judgement. It implies that any proper judgement cannot be contradictory, and truth has to be understood as being based on existence if anybody wants to understand this principle well.

Both logic and metaphysics study everything. However, logic studies things as known; metaphysics studies things as being. Therefore, logic is subordinated to metaphysics: to know is to understand what things are, and intelligence is the human faculty able to achieve that aim. Logic needs the previous concept of being. Truth is said of a judgement and refers to the correspondence between intelligence and reality. Such a correspondence with reality is not adjustable. There is no intermediate state between being and non-being in reality.

The principle of non-contradiction is firstly a judgement about reality: its deepest formulation is metaphysics (referring to the specific concepts of being and non-being), but there are many other possible formulations, e.g. psychological or logical. This principle is not an axiom or a mental postulate. It's the highest law of reality: beings are not contradictory. The logical derivation of this principle is a consequence from how intelligence knows reality (Alvira et al., 2001, pp. 44–45).

The principle of non-contradiction is achievable by essential induction, by which intelligence, assisted by experience, can perceive a necessary and universal link between a subject and a predicate (Sanguineti, 2007). All these terms (being, essence, induction, concept...) are specific in the Aristotelian-Thomistic philosophy. We can contradict ourselves when thinking or speaking. This happens when we move away from reality, because our reasoning is faulty. It is not possible to understand what is contradictory.

Other basic metaphysics principles based on the non-contradiction one are the principles of the 'excluded third' and 'identity'. The former says that there is no middle term between affirmation and negation; in other words, a thing is a being or a non-being without another alternative. Therefore, every proposition is necessarily true or false. The latter says that a thing is always what it is.

The principle of identity is explained by Forment as follows: "Having affirmed the concept of division, if this concept is denied in being, the concept of unity will be formed, that is, being is not divided in itself. What is one is an undivided being. Since unity is formally the negation of the internal division of being, it grounds the principle of identity immediately" (Forment, 2009, p. 276).

These principles are always assumed as fundamental knowledge for all sciences: they are not tested or discussed, but simply accepted and considered indispensable to reason. In contrast, metaphysics focuses its efforts on the deepest intellection of these concepts and primary judgements which grounds the activity of reasoning. Metaphysics aims at understanding reality from its ultimate and deepest causes.

In order to properly understand the metaphysical foundations of the concept of uncertainty, we need a more detailed analysis of the act of knowing and, in particular, of an accident of being, called 'relation'. D. Lindley highlights that statements of uncertainty are 'personalistic' (Lindley, 2014, p. 1). It is actually present in specific individuals as a mental state. Lindley

admits the possible existence of an inter-subjective agreement holding the same level of uncertainty in a particular question, such as, for instance, in the framework of some aspects of gambling. For our purposes, the crucial point is how Lindley states that there is usually a relation between the person who knows (or the group of people that expresses a common opinion or even a non-specific but singularised person) and that aspect of the world that is to be known and about which a statement is being made.

The metaphysical category 'relation' that Lindley highlights (in the 'Introduction' of his *Understanding Uncertainty*) is one of the so-called categories of being in the Aristotelian-Thomistic approach. There are ten of these categories: substance and nine accidents, one of them being relations. Aristotle's metaphysical concept of substance is linked to his conviction that it is possible for an individual to get to know the sensible world. In contrast, Plato thought that everything happening around us is in constant change. Aristotle held the opposite position because, after analysing the nature of change in the sensible world, he discovered what metaphysics calls the first division of being: what changes and the change itself. The first is stable and permanent, the substance, and the change relies on it. Only with the premise of the existence of a permanent element in beings shaping the sensible world, it is possible to maintain that there is a real knowledge or science about these beings and not a mere opinion, as Plato stated.

On the other hand, among the metaphysical accidents we find the relation. The accident is understood in metaphysics as a predicate of the substance. Relation refers substance towards something else: it is the arrangement or reference to another thing. Such 'reference to' is the key characteristic and what makes 'relation' different to the other accidents (Alvira et al., 2001, p. 72). There are two types of relation: (i) real, i.e. between real things; and (ii) logical, i.e. between concepts. Let us define certainty as the state of intelligence recognising the truth of its judgement (which becomes a statement). The mind assesses the adjustment between the concept obtained by intelligence and what the known object is in reality, that is, the logical truth of a specific judgment. This logical truth is bivalent: either there is an adjustment or there isn't. When there is an adjustment, the intelligence´s strength of assent can reach its highest possible degree: the degree of certainty. However, there may be certainty without truth, as we will later discuss.

The principles of Aristotelian-Thomistic general epistemology state that (Vernaux, 1971, pp. 124–129):

- Logical truth is bivalent: there are no degrees regarding its ultimate foundations on existence. Logical truth is indivisible: for a specific judgement there is no middle way between the adjustment of it to reality (or lack of it).
- Logical truth lies on the intelligence.
- Logical truth or error only exists when something is stated. Logical truth formally lies on the judgement of something being this or that (the union of concepts in judgement has an existential sense: the union is real). Thus, in the judgement there is reflexive dimension that perceives the conformity of the attribution to being.

In this framework, anything real, doubtful, rejected or imagined should be appraised in relation to existence. If we state that a chimera does not exist, this is a truth based on existence. In this doctoral dissertation, we will understand uncertainty from this Aristotelian-Thomistic standpoint, putting existence first.

1.2 Metaphysical theory of knowledge

1.2.1 The process of understanding

Immediately after, those basic ideas which in this matter are considered necessary to summarize are collected for the purposes of this thesis. They have been selected from a book written by De Torre (De Torre, 1980, pp. 160–170).

While the senses focus on sensible qualities, the intelligence wants to know what things are in themselves. The process of understanding includes three operations: (1) simple apprehension, whereby intelligence grasps the essence of things; (2) judgement, making statements (affirmations or negations); and (3) reasoning, drawing conclusions from those statements.

Intelligence, as potency, is actuated when it enters into contact with reality. Such contact with reality is reached by means of simple apprehension and judgement. Simple apprehension allows intelligence to produce concepts, and judgement to compose or divide concepts.

The sensibility of human beings is traditionally classified into external and internal senses. It is through the external senses that we are in direct contact with reality. Internal senses are not in direct contact with reality and they depend on the external senses for their supply. Four internal senses can be distinguished: consciousness (perception), memory, imagination and instincts. These internal senses are located between the external senses and the intelligence.

Instincts allow animals to react with emotion to sensible experiences – without thinking – considering them to be admissible or inadmissible. In human beings they are called 'cogitative power' or 'ratio particularis', in other words "reason applied to what is individual or particular", because it is the area of human knowledge wherein sensibility and intelligence overlap.

Once the senses are actuated by material reality, they produce the sensible experience, technically called 'phantasm', composed of the sensations of the external senses, and the perceptions, images and recollections of the internal senses. Intelligence is actuated by the 'phantasm'. Intelligence does not really start operating until the senses have reached a certain development.

Senses get to know sensible qualities and objects, but intelligence penetrates into the essence of things, which is purely intelligible. All the sensible qualities are accidents of things, since they do not exist by themselves. Thus we get to know the substance as the purely intelligible being that exists by itself.

Therefore, intelligence inquiries the intelligible essence of things, but it needs sensible experience to grasp their beings. Both the first concept of intelligence, being, and the first judgement, being is not non-being, are produced on the level of sensible experience.

De Torre points out that what is applied to the senses is also applied to the intelligence. There is not only receptivity in the process of knowledge, but also activity, since intelligence acts on the products of the senses by enlightening them in such a way that the intelligible essence they have is shown.

The process whereby intelligence goes from the sensible to the intelligible is called ‘abstraction’. The content of the intelligible forms that we grasp through the senses has to be ‘separated’ from the individual matter. This active role of the intelligence is called ‘agent intellect’.

The singular material essences, because of being material, are not suitable to be understood by the human intelligence, while the universal essences, because of being immaterial, are intelligible. However, although the former are potentially intelligible, they cannot become so by themselves because they need another substance to actuate them, separating them from the matter. This is what justifies the agent intellect, a faculty of the intelligence that enables explanation of how the universal intelligible concepts are formed. The human mind becomes an intellectual faculty in act by the in-formation coming from what is intelligible. This intelligibility is first a potential property in sensible things. Then, our intelligence is actuated in the process of abstraction by which we grasp the intelligibility of sensible things.

The agent intellect is like a light that exposes the intelligible form, but the so-called ‘receptive’ or ‘possible intellect’ is necessary to carry out the act of understanding by the intelligence. The enlightened intelligible forms are impressed in the receptive or possible intellect (the impressed forms are called ‘intelligible species’), being stored in the intellectual memory. Intelligible species are also called ‘intentional species’.

Therefore, the act of understanding is the act of two potencies: (1) the potency of the thing to be known by the intelligence, which is called ‘passive potency’, and (2) the potency of the intelligence to know the thing, which is called ‘active potency’.

Intelligence is initially a passive faculty and needs to be actuated by several intelligible species or forms in order to reach eventually its object. No single intentional species gives us the entire object. Intelligence needs a process whereby it has to be ‘in-formed’ by many species, all of them being related to the same thing. This radical dependency of the intelligence on reality is the kernel of the realist theory of knowledge.

The intellectual or impressed species whereby the possible intellect is actuated enables intelligence to express a mental word which is called ‘concept’ or ‘expressed species’. Both impressed and expressed species are the medium for intelligence to grasp reality. It is the classic example of the external glass of a house and the mirror inside a room wherein something outside the house is reflected whose image comes to it through the aforementioned external glass. The impressed species is similar to the external glass and the expressed species to the reflected image. In any case, both species are signs of the reality that they signify. De Torre expresses this process by saying that reality actuates intelligence, fertilises it somehow and the fruit is the concept.

The concept is different from the sensible image. The sensible image is a sensible, individual and material accident. The concept is an intelligible, universal and immaterial reality that intelligence grasps from sensible images.

However, the sensist theories of knowledge reduce the objects of human knowledge to merely more or less complex sensations. Therefore, the agent intellect loses its raison d’être. If the universal corporeal essences are not knowable, the existence of the agent intellect that is responsible for that function makes no sense. In those theories the universal concepts are

reduced to pure images and, in this way, it is not possible to explain how the properties of certain objects of knowledge called 'universals' can be attributed to a plurality of things.

On the other hand, the intellectualist theories of knowledge disregard the mediation of the sensible knowledge. They believe that the human mind is able, by itself, to grasp the universals directly.

Once the impressed species actuates the passive understanding, this one carries out the knowledge operation: the form becomes manifest to the mind, with the impressed species being a resemblance of the form of the known real thing. What is finally known is the form, not the impressed species. This form is what determines the mind. We cannot grasp the form from the impressed species in an immediate and spontaneous way.

1.2.2 Act of knowledge and intentional object

To know is a perfect act because it owns its aim: "we see and we have seen, …, we think and we have thought," as Aristotle says in his treatise on his First Philosophy (Aristotle, Metaphysics, IX, 6, 1048b 18–35). As Corazon explains: "in order to know, therefore, we do not have to do another thing previously, we have not to use means or to carry out actions, or transitive movements, but we know directly and immediately," and this author emphasises: "what Aristotle has described is the act of knowing in itself, not what the person who is going to do it has to do to get it. That's why he also distinguishes between to think and to learn; to learn requires time and it might not reach its aim. To think, on the contrary, is a perfect act" (Corazon, 2002, pp. 48–49).

To know is, therefore, to own; that is, we cannot say that we have seen and we have not seen at the same time, that we have known something and we have not known something at the same time; that is, as Corazon underlines: "there cannot be act without object, and object without act" (Corazon, 2002, p. 49). Object is not a synonym of thing, but what comes to the mind of the person who knows, what he/she has opposite him/her (ob-iectum). Llano clarifies: "Objectivity is the gnoseological status reached by known things as known according to their own realities. It is not, therefore, a real dimension" (Llano, 1991, pp. 36–37).

The apprehension belonging to the act of knowing is intentional; that is, what is apprehended is the object. Insofar as the object is apprehended, something is known. Aristotle expresses well that the thought object is not another thing than the act of understanding in itself: "what is understood in act is the intelligence acting" (Aristotle, De Anima, III, 7, 431a 1 s). In saying this he asserted that to know in act is to know something, not that what is known is the own act of knowing. Corazon underlines: "acts are not transitive actions, the mind does not need to build or make up what is known, but to know (in act) is the same as what is known" (Corazon, 2002, p. 50).

Corazon clarifies the following: "The object is a pure sign, something that refers to another thing without signifier (a pure significance without signifier)… If the object was made up of signifier and significance, an infinite open process would happen, because the signifier would be arbitrary and the significance would be taken from an earlier stage, unless with nominalism it is asserted that ideas (intuitions) are 'natural' signs." Later on he says: "The signifier is the own act of understanding, that is, what is known in act. When we know, we know realities, not ideas. What we call ideas are the formal signs which we are speaking of. To know the own

ideas requires us to focus our attention on them, regardless of their own significance" (Corazon, 2002, p. 50).

In other words, the lack of an intermediate stage between the act of knowing and what is known is an essential aspect of intentionality. What come to the mind are not ideas or images but what is known. Since knowing is a perfect act, when intelligence knows something, it reaches its aim simultaneously.

Corazon puts his finger on the wound when he says: "Knowledge is often understood in modern philosophy as a relationship between subject and object, but it is wrong. He who knows is not the subject of the act of knowing immediately, but the faculty: sight, hearing, intelligence, and so on. It is a big mistake to reduce a person to a subject. We are not a *res cogitans* as Descartes says, because a person cannot be reduced to the act of knowing, and because a person does not appear in such an act. To own the aim means that between the act and the object there is nothing in an intermediate stage, because in this case what would be owned would be such an intermediate thing" (Corazon, 2002, p. 51).

Therefore, "The terms 'idea', 'species', 'representation', and so on, are all ways of calling the owned object by the faculty (of knowledge). It is a mistake to convert them into things or substances. Ideas are not things in the mind, but the intelligence in act. As these acts are multiple, none of them either equals or exhausts the faculty…" (Corazon, 2002, p. 51, fn. 9).

The intentionality of knowledge is lost in the Cartesian philosophy because what is known is not the reality but ideas. The step towards reality is made through the principle of causality, indirectly. As a consequence of the above-mentioned, we cannot say that we can know the reality directly, that is, to know does not have to do with to learn what things are, but to reach an idea useful to live. For Kant, objects per se remain unknown because we are only able to perceive intuitions, mere phenomenical representations.

The modern theories of knowledge turn truth into a product of reason, set science above metaphysics, and set knowledge depending on vital necessities or arbitrary decisions of the will.

Corazon makes a clever and deep criticism of 'subjectivism' that he equals to 'voluntarism'. He asserts: "Before knowledge an attitude of suspect is always possible, to doubt about its truth. When that attitude is spread to all that is known, the will takes charge of controlling all the contents of awareness, which have to pass an exam, a clarification, to check if they satisfy or not the desire for accuracy demanded by the subject as a prior requirement for its admission," and concludes outright: "At the root of subjective attitudes there is a desire for accuracy before knowledge. First is not to know something but to be sure, to avoid mistakes. If, as we have seen before, what is known in act is the intelligence acting, then we have to say that he/she who adopts such an attitude has resigned to know, and he/she settles for being sure that the object (not reality) does not hide anything not previously examined, is not confusing, and does not disappoint the expectations of the subject. The subjective attitudes set the will before intelligence: they do not base certainty on evidence but rather they consider what the subject admits as certain to be evident" (Corazon, 2002, pp. 56–57).

Such an intended subjective attitude is not real because it is not possible to anticipate knowledge, to reach an earlier knowledge to all knowledge, knowledge of the knowledge in itself before getting to know. To reflect, to think about thinking, is always something posterior,

because what is immediate is to know something. That way the human being is a being open to reality immediately, he/she does not open to it by means of a voluntary decision. His/her act of understanding goes before every voluntary act.

Corazon concludes by saying the following: "The metaphysical realism is based on the gnoseological realism: what is known immediately is the reality, not the knowledge in itself, which hides to reveal what is known" (Corazon, 2002, p. 58).

1.2.3 Knowledge criticism

Metaphysics can explain the meaning of knowledge as showing or expressing being before our intellectual capacity. It considers that intellectual faculty is suitable for being actuated by any being.

There are two essential elements in every act of understanding: what is known (being) and who knows (the human mind). Both terms are understood in their more universal dimension: being as something common to all of them, and the mind as actuatable by any of them.

The key question that constitutes the kernel of what is nowadays called 'knowledge criticism' is the following: is our mind actuated by the true reality of things? If so, how can we know it?

The intellectual faculty is able to take charge of the sensory knowledge and, besides, to think about its own acts of understanding. All that proves that the human mind is able to reach its own goal, though its judgements are not infallible. If it were supported that the human mind cannot reach any truth and certainty, it would make no sense to use it to measure their true value.

When we wonder whether the mind has truly understood what things are in reality (whether it has been actuated by the being of things) we care about the concept of logic truth. This concept of truth is different from the concept of ontological truth, which is a property of the being: its intelligibility.

The logic truth is defined as the agreement of the mind with the reality that it can apprehend. It is a con-formation. When the judgement made by the mind is true, it acquires the same form as the thing has in itself. It is a formal or intentional identification. The mind and what is known are 'intentionally' made the same thing. And, in addition, the agreement that defines the logic truth is a known one. The agreement between a concept and its extra-mental substance is only knowable in a rigorous and effective way through the operation called 'judgement'. Making a judgement is how the agreement is formally expressed. It is said that the agreement is known when besides knowing the terms of a proposition the special relationship between them is apprehended by the verb 'to be', which implies, in our mind, the act of making a judgement. That's why the logic truth is a property of a judgement in the sense that only in a judgment this kind of truth can occur.

According to Sanguineti, the above-mentioned arguments can be summarized by saying that the basic concepts are made by abstraction from the sensible knowledge (Sanguineti, 2007, p. 36). Intelligence gets to grasp step by step the intelligible aspect of things from the experience data. The science that studies in detail how concepts are made (basically through reflection and construction) is called 'psychology'. Concepts are abstract because they mean an essence abstracted from singulars that own it.

To say that concepts are abstract does not mean to say that singulars are unknowable. We can know them as long as our intelligence, in the process of making concepts, comes back to the sensible experience (conversio ad phantasmata) and understands concepts realised in singulars.

To abstract is to consider some aspects of things, regardless of others, that in reality go together. All concepts are abstract in the sense that they leave the singular, from which they have abstracted the essence meant in those notions. When we refer a universal concept to a singular we are stating its essence. Singulars are the unique existing beings, and essences express what intelligence is able to grasp as common to all of them.

Nominalism denies the existence of common essences in things, denies the real existence of universal concepts. Only individuals and individual properties are in existence for nominalists. They say that the common words used in normal language are justified by mental economy, by the impossibility of naming each thing differently. They are only useful for classifying more or less similar objects. They think that such similarities are relative and repetitive facts in the past, and there is no reason why they have to happen in the future. Therefore, only terms are universal for them, because they can be applied to many things.

However, common words really express universal concepts. Words are signs of acts of understanding and a content that is understood. Concepts are immaterial, but this does not mean that they are something vague because they have a precise mental content. They mean real natures: what is thought by concepts is real in singulars. When we wonder what something is, we do not wonder what such a thing should be named, though the answer to the first question also answers the second one. If names were not to do with what things are, they would be reduced merely to what we think or are able to do with them.

Being the mind able to grasp what something is, the essence of what the object of an act of understanding is, what the philosophy of being underlines is that this is possible starting from the sensible qualities of the known object. It does not mean that the essence is perfectly understood from the first sensible experiences. Multiple experiences are usually needed to understand what is perceived. So, to understand something is the act of knowledge that allows us to grasp its essence. When we name things significantly, we make acts of knowledge of that kind.

However, though the language allows us to attribute the essence of something to something real, it is not the essence what is sensorially perceived. The philosophy of being points out that it is necessary to abstract the intelligible essence of material beings to apprehend it from the individual determinations, which are accidental and extrinsic to it, but that they are with it in the beings.

From the Aristotle-Thomas metaphysics of knowledge it can be understood that induction is a way of reasoning that is able to reach a universal conclusion from singular propositions or partial data. We can reach this universal conclusion from the individuals that make up a universal concept, which all of them share. This way, it is possible to reach the intelligible level, that of conceptual and universal truths, starting from the sensible and material world. This is what Aristotle supported in explaining how the mind understands what is intelligible. He called it going from sensory to intelligibility induction. This transit is basic to form concepts in the mind and, therefore, to reach universal and necessary truths.

On the other hand, St. Thomas Aquinas states that two acts of reason have to be considered in the judgement: that which allows us to apprehend the truth of something, and that which assents to what has been apprehended. Corazon distinguishes, as a consequence of the assertion of St. Thomas, between 'the content of the proposition' and 'the assertive strength' (Corazon, 2002, p. 193). He does so because he is interested in underlining, in this context, that the assent is led by the intelligence, not by the will, to avoid voluntarism. Different mental states are possible before a proposition: certainty, doubt, opinion and faith, and each one is characterised by different assertive strength on the truthfulness of the content of the proposition. Nevertheless, he underlines that the voluntary decision to accept the truth is indispensable to act, because the will can also act regardless of the truth the intelligence perceives in a proposition.

Truth, certainty, doubt, evidence and so on can only happen in judgement, never in concepts. It is not true that there are, on the one hand, clear and distinct ideas and, on the other hand, unclear and obscure ideas; it is another thing that he/she who knows wants to know more, but each idea is what it is, and it is known or not known: there is not an intermediate state. With regard to assessment, evidence in judgements is always referred to its dependence on 'principles'; the 'primary principles' are evident in themselves, and if a judgement is made from those principles, it is evident. If the 'principles' belong to any science, they are hypothetical (by default), and the conclusions, though rigorous from the logic viewpoint, are as hypothetical as the principles. When we assert something without the support of the principles, the conclusions cannot be certain, and that's why intelligence, by itself, cannot assent to the conclusion unless the will inclines it to do that. In this case, the assertive strength is not entirely given by intelligence, though it is only the intelligence that perceives the degree of correspondence between what is thought and what is real.

1.3 Principle of causality

The Aristotle-Thomas approach understands the principle of causality as the dynamics of being: the mutual influence between beings by means of causation. The being of things allows them to act and, therefore, they can influence the being of other things acting upon them or their becoming.

The notions of cause and effect are common to all the ordinary experience of all human beings, basically regarding what is called 'efficient causality': external, internal and internal-external (see Lucena, 2016, p. 153).

The above-mentioned experience of causality is evident for human beings. It makes no sense to show its existence because it always ends up imposing on the intelligence by itself. The foundation of that reality is being: as it is, it can cause. Likewise, it makes no sense, from the philosophy of being, to study the principle of causality without suitably understanding the notion of being first. It is not possible to understand causality without knowing some beings, because causality is said of things that we call 'causes', and affect other things that we call 'effects'.

On the other hand, the notions of cause and effect are mutually needed. They are apprehended by their mutual linking. That which is a cause is a cause of something, and the effect entails a causal origin. The effect is always dependent on the cause: that relationship has to be real. However, the relationship between the cause and the effect may only be rational

when the concept of cause is based on being (for example, the being of God with respect to that of creatures). The cause-effect interdependence between things is called the 'principle of causality'.

According to Millan, "the most proper and rigorous meaning of the term cause is that established by consecutive determinations of the notion 'principle'" (Millan, 2002, p. 75). See (Lucena, 2016, pp. 153–155, for further details).

Hume considered the concept of cause to be based on a temporal succession of phenomena. This idea is linked to his empiricism, in which reality is exclusively related to data of experience. In the notion of cause of the philosophy of being there is a subordination of the effect to its cause in the order of being or becoming, but not unavoidably in the temporal order as Hume holds. This statement is compatible with considering such a metaphysical subordination to be also temporal in the scope of empirical sciences. The temporal succession is, for Hume, the key to the cause-effect relationship.

Cause is not always data of experience. Usually it has to be inferred by reasoning from empirical data. Moreover, sometimes, in spite of the data and the subsequent inferences, the cause of the observed effect does not get to be known. That it does not mean that a true cause does not exist.

Hume denies validity to the concept of cause because, in his opinion, experience cannot confirm it. What we call cause is for him a mere sensorial impression from which we expect another impression to follow. The ground for this expectation is the force of habit. There is no logic basis for Hume for such an expectation. He emphasises that experience cannot guarantee that what is observed will happen again in the future.

Millan distinguishes three aspects in Hume's thought about the concept of cause (Millan, 2002, pp. 79–80):

a) The idea of cause in itself:

   In common language the term 'cause' is not only used to mean sensible impressions. The idea of cause is not only thought of as a current or possible empirical datum. Besides, and this is decisive, Millan underlines that the general idea of cause cannot be considered an effective empirical datum. It is possible to perceive a thing sensorially, which is a cause, but not that such a thing is a cause, the same way that being-body is not sensorially perceptible, though those beings that have bodies are perceptible. As an empiricist, Hume is strictly nominalist.

   Words are common or universal because of their meanings, not because of their own beings; they are merely semantic universals. An act of understanding whose object is any individual belonging to a determined class is not universal in itself. This is what is done when a universal meaning is associated with a word. Every act, and the fact of associating a universal meaning with a word, is something individual in itself. What is universal in an act of understanding is not the act in itself, but what it represents: something abstract that does not fit for only one specific individual belonging to a determined class, but universally, for all its members.

The difference between semantic and representative universals is that the former are merely words, and the latter mental acts. The philosophy of being points out that there are no beings in existence that, in themselves, are not individuals.

According to Hume, general ideas or images formed in the mind do not come from particular ones. There is no induction from what is particular to what is universal because, in fact, all ideas are particular for him. These latter become general or universal through the evocation in the mind of other similar particular ideas. General ideas come from association of ideas. They have a merely psychological link, and they are called a common name disregarding any difference. Thanks to such an association, evocation or memory, custom or habit, when an object is present before mind, its individual image receives this common name.

b) The genesis of the idea of cause:

The being of the cause, though sensorially imperceptible, can intellectively be grasped by an abstraction whose starting point is given by certain simple and well-known facts regarding our experience, as was described at the beginning of this section. The idea of cause is obtained by abstraction from those facts. However, we have no experience of *ab extra ad extra* causality, the influence of beings external to us over other beings external in the same way.

c) The question of its application:

If Hume attributes to the force of habit the expectation of sequential sensory impressions, it will have to be asserted that the habit is its cause. So, Hume is begging the question: habit explains causality and causality is explained by habits. Millan underlines that Hume admits in this way an *ab intra ad intra* causality.

The philosophy of being shows the principle of causality through a deductive argument starting from the idea of 'becoming': non existing entitities that come later to exist owe their existence to some other being. That is, they have been caused to exist.

The need for being caused is logically inferred from the conjunction of its fact of being and the possibility of not being, which is inferred, in turn, from the fact of not having been. The deductive reasoning that allows the principle of causality to be shown is the following:

What becomes may not be, because it was nothing.
However, what may not be, and nonetheless is, is not by itself, since, if it had been by itself, it would not have lacked what is necessary to be;
Therefore, what becomes is dependent on another, that is, is caused.

In this reasoning there is a mediate inference called a 'syllogism'. In order to get a better understanding of how the syllogism is made following their logic rules, it suits us to reorganise what has been expressed before in the following new way:

The minor term of the syllogism is the statement "What becomes" (A), and the major term "to be dependent on another" (C). The conclusion can be formulated like this to make it clearer that it is formally a proposition: "What becomes is dependent on another" (A is C).

The middle term links both premises, the major and minor ones, and it is the following: "What may not be" (B).

The reasoning is a simple syllogism, with the following structure:

A is B: "What becomes may not be"
B is C: "What may not be is dependent on another"
A is C: "What becomes is dependent on another"

What is proved with this syllogism is not that everything is caused, or that every being is caused, because the concept of 'what becomes' does not fit to mean being or every being, that is, being is not equal to being caused. 'What becomes' is the same as 'what is caused'. And the concept of being is compatible with the possibility of being by itself.

Hume's concept of *causa sui* is contradictory. In order to be cause of itself it is necessary to be and, at the same time, not to be – the former to be able to cause something and the latter to be something caused. And, on the other hand, the concept of necessary being is not contradictory because there is no reason at all to accept that all existence should be contingent. If Hume thinks in this way it is because he restricted reality to what is object of experience. To exist is not the same as to exist without necessity, or to exist without cause. What exists without cause exists by itself, that is, what exists necessarily.

Millan emphasises that the use of the principle of causality, when exterior beings are considered, is not an ontological issue but methodological, which is solved, when there is enough experience, by applying the laws of induction, and taking into account all kinds of necessary conditions in order not to proceed without foundation. And in those cases wherein there was no basis to make an induction rightly, the principle of causality would continue to be valid. 'To be an effect' is not the same as 'to be known as an effect', likewise 'to be' is not the same as 'to be known'.

Millan asserts that the concept of cause as envisioned by Kant overcomes that of empiricists and positivists for two reasons clearly expressed in his *Critique of Pure Reason*:

a) Because he is aware of the supra-empirical reality of the concept.
b) Because he includes the concept of subordination between the cause and the effect, and, therefore, he does not accept the reductionism of the mere temporal succession.

But the concept of cause is, for Kant, no more than a mere form of the mind. That concept has no other function for him than to be useful for linking mere phenomena. Instead of referring to something real, it refers phenomena each one to the other. It can be applied to order phenomena in their temporal dimension. Thus, the value of the principle of causality is inter-phenomenical, and presupposes time as an indispensable condition. The concept of cause is not useful to understand what things are in themselves, but to order data from experience restricted to offer us the results of the action of things on our sensibility.

Kant does not accept that the concept of cause can be abstracted from some empirical data. The denial of this possibility, as happened with Hume, has to lead forcefully to scepticism. Millan states that there is no philosophical argument to contradict the value of an immediate datum obtained from our internal experience.

The philosophy of being formulates this principle as follows: “what is convenient to something and does not belong to its essence, belongs to it for some cause” (St. Thomas Aquinas, *Summa contra gentiles*, I, 22). Perfections are observable in every being as parts of its essence. However, there are also perfections that are not parts of it, and these have to be necessarily caused by a different agent from the considered being. All beings agree on being, which does not belong to their essences. Therefore, it implies that the being of beings demands a really different cause. Essence is a principle of diversity, while being is a principle of unity. This reasoning is the basis of the fourth way of demonstrating the existence of God written by St. Thomas.

The notions of cause and effect include the notion of being, and this does not imply either to be caused or to cause. It is not logically consistent to assert that something is caused for itself. The principle of causality is not deduced from the notion of being, but it is inductively discovered as our experience warns us of the limitation and finite quality of any effect.

Arana tries to clarify, in his work titled *The Basements of the Universe*, the concept of cause following the same defining criterion as Aristotle, confining himself mainly to the use of that term in common language (Arana, 2012, pp. 99–101). That leads him to say that in any of the classic meanings of cause there is a relationship between two terms: the subject of the causal action, simply called ‘cause’, and the object of the causal action, called ‘effect’. And Arana underlines that we run into a first ambiguity consisting of using the same term for the subject of the causal action and the action itself, and a second one because sometimes the effect points out to the result of the causal action, and others to the passive subject of the causal process.

Arana proposes an alternative explanation of the concept of cause focusing on the relationship between both terms: the agent and patient subjects of the causal action, the whole thing regardless of the nature of terms such as things, processes, phenomena, concepts, etc., and the type of relationship (real, ideal, material, formal, etc.).

Arana says (Arana, 2012, p. 100) that the causal relationship tends to nominalise itself, whereby he proposes the following explanatory structure of the elements that make up causality:

Agent subject (1) – causal action (2) – effect (3) – patient subject (4)

The word ‘cause’ refers mainly to (2), though by extension to (1) as well. The word ‘effect’ refers mainly to (3), the result of the causal action, though to (4) as well.

This structure, Arana continues, is useful to clarify, partially, Aristotle’s doctrine of the four causes:

a) The material cause has to do, above all, with the patient subject (4).
b) The efficient cause refers to the agent subject (1), and as similar to the form, to the causal action (2).
c) The formal cause covers both the causal action (2) and the effect (3).
d) The final cause refers to the effect (3) as predicted and desired by the agent subject.

In addition, the meaning of the structure can change according to the nature of the terms of the relationship, in other words, when it is applied to processes instead of things, for example.

In this case the patient agent (4) is modified because it stops being a substance or a static being to turn into a dynamic sequence with a higher degree of formality than in another case. Arana illustrates this warning that the matter-form relationship is relative: a piece of bronze can be considered a form with respect to the basic elements that it comprises (tin and copper), but also as a matter when referring to objects manufactured with it (bells, statues, cannons, etc.).

What Arana underlines with the earlier example is the impossibility of universally extending the meanings of Aristotle's causality, and any sort of doctrine directly or indirectly based on it. He asserts: "Cause is probably the most versatile among all kinds of categories of determination, both because of the amount of senses given by language, and the relativity of the hylomorphic theory, which is the foundation of its older and more long-term theoretical development. This is the reason why it has been able to adapt to the human mind evolution (we continue to speak about 'causes' as often as before), but also it has been rejected by those who demanded precision and univocity in scientific terms and concepts" (Arana, 2012, pp. 100–101).

Arana points out that modern science has turned to laws – excluding the concept of cause – into the basic instrument of his work for the following reasons (Arana, 2012, p. 104):

a) Because it is a less ambiguous concept than that of cause.
b) Because, in spite of it, it continues to be very versatile.
c) Because it has found a way of being expressed with all the required precision by means of mathematics and the fact of being included in convincing demonstrative reasoning.

Arana, however, states the following: "It is not only reasonable to interpret the physic-mathematical law in the same way as formal causality. Other sorts of causes can also be related to that, because it does not consist only of a gaunt structure, but those magnitudes that allow the law to be applied to a great amount of particular cases appear as variables in its formulation. When replacing these variables with the values according to the studied situation it is fair to say that the material cause has been included in a formulation that has lost in generality what it has gained in precision. In turn, in functional expressions that include time as independent variable, the starting values of variables could be interpreted, together with the dynamic specificity of the law by itself, as the efficient cause of the transformations that happened … Lastly, if natural laws by themselves and the absolute initial conditions of the universe could be considered as the chosen object regarding the results that they are going to produce, the final cause would appear as if it were the most natural thing in the legalistic interpretation of the world" (Arana, 2012, pp. 106–107).

1.4 Metaphysics and logic

While metaphysics studies reality, logic studies relations among beings in the human mind. That is why metaphysics should be studied before logic, and so it was for centuries in many philosophy teaching centres. There has always been a trend to confuse the order of logic with that of metaphysics, the order of ideas with that of real things. If a person has a proper training in metaphysics, he/she will never maintain something as trite as "love is the cause of marriage" because he/she will know that love is, metaphysically speaking, an accident, the activity of a human being, the so-called 'action accident'. Only people, who are substance, can cause marriage by means of their love and conscious, deliberate and mutual acceptance. This

is an example of the common confusion between what exists by itself in reality – substance – and what exists by itself only in the mind, as concepts: accidents (De Torre, 1980, pp. 57–58).

This approach would allow us to appraise in a completely different way the debate about the nature of uncertainty, whether it actually exists in the world or just in our minds. From the Aristotle-Thomas perspective, uncertainty is the outcome of the intellectual activity of a human being. It is a direct consequence of our intelligence, this latter being an active faculty (an accident) of our mind. The operation of intelligence (sometimes) brings about uncertainty. Uncertainty is thus an accident of intelligence. For a realist philosopher in the Aristotle-Thomas tradition, uncertainty has a mental being. Uncertainty and indeterminacy are not the same concept, and both concepts will be explained in detail in Section 4.1.5.

The following can be read in a classic treatise of logic from a Spanish university: "Only a metaphysical philosophy can justify the possibility of a scientific knowledge, and the validity of methods of different sciences. Evading any conviction of truth, either it is inconsistent with the real scientific task, or it takes to a scepticism leading to destroying any scientific motivation" (Sanguineti, 2007, p. 231). If those concepts on which we want to build a rational foundation of an aspect of the reality are not based on the first philosophy, such an aspiration will not be possible. If that assertion has been made with such firmness it is because the capacity of human beings to know, truly, reality through their knowledge faculties (senses, intelligence, memory, and so on) is considered to be evident under Aristotle's philosophy of being, although in a limited way. Such a certainty is metaphysical and is based on the first principles of that science.

Einstein once declared in relation to these matters: "Without the belief that human beings are able to understand reality by means of theoretical constructions, without the belief of an internal harmony of our world, science would not be possible. This belief is and will always be the main reason for any scientific creation. In all our efforts, in each dramatic struggle between old and new conceptions, we acknowledge the aspiration to understand, the always firm belief in the harmony of our world, continuously reasserted against the obstacles opposed to our comprehension" (Einstein, 1978, p. 276).

The Spanish philosopher De Torre points out in his work titled *Christian Philosophy* the relevance of St. Thomas Aquinas's explanation about the difference between philosophy and the other sciences according to different degrees of elevation above matter (De Torre, 1980, pp. 44–45). 'Matter' is the word used by Aquinas to designate the world perceived by our senses, and the first characteristic of this world is that it is in continuous change. But science hopes to become a stable, true, permanent, valid and accurate knowledge. Since matter is in continuous change, scientific knowledge has to rise above matter. Experimental sciences try to discover patterns and universal and stable laws in the corporeal world. At this level of knowledge of reality it is not so interesting to study characteristics of isolated individuals (usually variable) as those of species (universal and permanent).

A second degree of elevation above matter is that related to the study of beings, which, although they cannot exist without matter, can be thought about without matter. These are abstract quantities or dimensions, both continuous (lines, surfaces, volumes) and discrete (numbers). Quantity, as an accident, cannot exist without matter, but it can be thought about without matter. The science of abstract quantities is called mathematics. As this science is placed at a higher level of abstraction above matter than experimental sciences, it has more certain and accurate postulates and laws.

Lastly, the third degree of abstraction above matter studies both beings that can exist without matter and beings thought about without matter. It is called the 'transcendental level of knowledge', the level of being. At this level, the whole reality can be studied, as was mentioned before, including that related to beings that are not material by nature (such as God, the human soul, or spiritual beings such as angels).

Serious doubts were cast upon the philosophy of being as early as the 14th century when the so-called (by historians) Modern Age started. The rejection was closely bound to the crisis lived in Europe within Christendom and the development of the philosophy in the course of that period of time. One of the key points in understanding the above-mentioned crisis was the rejection of the dependence of human reason on being, i.e. on reality outside reason. Philosophy turns into immanentism: all reality is reduced to, and centred on, man. This philosophical conception leads, on the one hand, to liberalism, and later to anarchism, and on the other hand to socialism, leading to communism later on. Immanentism (*immanere* = to remain in) can be summarised by saying that nothing transcends man: reality is a creation of the human mind. Philosophy begins in human consciousness and there it remains.

The current prevailing culture is deliberately anti-metaphysical[2]. Making human thought the centre of reality is the essence of modern philosophy; it tries to subjectify reality (De Torre, 1980, p. 290). Protestantism had subjectified the Christian faith, and Descartes subjectified reason. The Aristotle-Thomas realism has always underlined that from an 'idea' one cannot prove the existence of anything in reality. However, Descartes, for example, tried 'to prove' the existence of the world outside himself by means of the idea of cause, considered by him to be innate, inverting the natural order of things by trying to prove what is more evident (the outside world) through what is less evident (an idea in the mind).

After Descartes, there were two schools of thought: rationalism and empiricism. Both schools stemmed from the Cartesian principle of immanence. Later on an outstanding figure in the history of philosophy appeared: Kant. He attempted a synthesis of the above-mentioned philosophical schools. After an enormous effort, he concluded that reality rotates around the mind. Metaphysics has no basis as a science for him (on the one hand he states that it does not have sensible content, and on the other hand that the concept of cause is only valid when applied to observable phenomena), it is only a 'natural disposition' of our mind trying to unify all experience.

Philosophies after Kant (the Romantic Age: the idealism of Hegel, Marxism, positivism, pragmatism and evolutionism) denied again and again the validity of reason in the field of metaphysics. Existentialism was not an exception.

Lastly, this introduction would be unfinished, failing to mention that human language, the transmitter of knowledge reached by the activities of the intelligence (concepts, judgements, syllogisms, etc.) and the activities of the will (aims, orders, feelings, desires, etc.), has a deeply rooted logic in metaphysics: its direct relationship with the being of things. Human beings name things because they are able to know them, and they can know them because they are: names reflect the being of things, though they also suggest the way by which human beings know them (Sanguineti, 2007, p. 64).

[2] Extracted from the homily given by Pope John Paul II in the Shrine of Nettuno (Italy) on 1st September 1979.

The clear choice of the Aristotle-Thomas realism as the theoretical foundation of this work is justified in the belief that, in the opinion of this author, there is not a more solid rational base to carry it out.

All the above-mentioned comments have much to do with the terminology used in the conclusions of forensic reports because there is a term with a special theoretical and practical importance: the concept of probability. Experimental science is essentially inductive and it may only achieve physical certainties, as will be shown in detail later. Therefore, conclusions will be probabilistic. Probability as a measurement of uncertainty is an essential concept in evaluating the strength of scientific evidence to clear up crimes. If uncertainty is defined from the philosophy of being, from this position – in our opinion – the different proposals about the concept of probability made by the different predominant schools throughout history are illuminated.

1.5 Metaphysics and science

In his work titled *Philosophy of Experimental Science*, Artigas tackles the philosophical impact of science. He carries out an excellent summarising exercise in pretended dialogue with J. Monod, H. Marcuse, E. Agazzi, K. Hübner, K. Popper, M. Wartofsky, G. Radnitzky, J. Habermas, I. Johansson, S. Jaki, T. Settle, T. Khun, S. Toulmin, I. Lakatos, W. Stegmüller, P. Feyerabend and H. Skolimowsky: (1) the sense of the scientific activity; (2) the ontology of science; (3) science and rationality; (4) the problem of demarcation; (5) scientificism; (6) the reliability of science; and a final discussion about the state of the art in philosophy of science (Artigas, 1992, pp. 363–419).

With regard to (1) he remarks, in contrast to Monod's thesis about scientific objectivity: "Certainly, the quantitative and experimental study of nature discards the explicit consideration of final causes; but it supposes the existence of a natural order, and discovers the laws which form it progressively, which is equal to asserting the existence in nature of an intrinsic aim." And in contrast to Monod's argument about final causes, he says the following: "Surely, it is not possible to prove the non-existence of final causes. Experimental science studies nature, so its method prevents it from thinking about what is found outside, as in the case of God, or the divine plans …, because they are transcendent final causes that go beyond the natural order." However, Artigas underlines that "the case is different dealing with immanent final causes, that is, with the aim of natural agents, because the natural order unveiled by science implies the existence of immanent finalities, and this is especially manifest at the level of living beings …"

Artigas, mentioning Agazzi in his work titled *Topics and Problems of Philosophy of Physics*, states that experimental science cannot tackle the study of reality in itself as a whole, a perspective corresponding to philosophy. Underneath Monod's thesis he uncovers scientificism, because scientific knowledge is the only one which he considers to be valid. He rejects the notion that the rationality of experimental science can be reduced to an analytical level, as Monod or Marcuse defends due to their scientificism, because "it discards essential aspects of the scientific method". For Artigas, "experimental science is based on realistic ontological and gnoseological foundations, without which it would lack sense", and this task, he continues, "is properly philosophical and involves a perspective of totality that goes beyond partial objectifications, which are typical of experimental science".

As regards (2), Artigas stresses that “When it is asserted that science implies a materialist, instrumentalist, or reductionist ontology, or of any other type, we are not really talking about experimental science in itself, but about scientificist opinions that extrapolate partial aspects of science, until turning them into a philosophical view which cannot be properly justified based on science”. In contrast to Hübner, who agrees with the above-mentioned viewpoint though he offers a historicist explanation, he supports the idea that the ontological foundations on which experimental science is based have their own value and they are non-dependent on historical situations. Artigas emphasizes: “Behind all these problems we find the difficulty of explaining the real value of scientific knowledge.”

With regard to (3), Artigas points out that the biggest epistemological disagreements happen in the relationship between methodology and gnoseology, that is, how science reaches its knowledgeable goal. According to M. Artigas, “neopositivism adopted an empiricist philosophy in which the value of knowledge was dependent on immediate perceptions”; the result was a methodology incapable of explaining the realistic scope of knowledge and, in addition, did not match the methods actually used in research. In those circumstances, inadequate rational reconstructions were used. The debate on rationality focused more on epistemology of a logic type proposed by Popper and the specifically sociological analysis of Khun. Both perspectives point out actual aspects of the scientific method, but they are extrapolated in such a way that they do not explain actual procedures and the value of knowledge. The proposed later attempts to harmonise both viewpoints, as in the case of Lakatos and Stegmüller, have contributed both to clarifying particular aspects and to leading to new problems because they have not used a consistent realistic gnoseology either.”

In terms of (4), Artigas says that Popper is mainly responsible for considering that problem to be the most important one of the theory of knowledge in our time. Artigas points out that Popper establishes the criterion of demarcation, conditioning it to an ambiguous conception of metaphysics in his works. Artigas presents the following idea regarding what Popper says in his works: “But, how is it possible to refer to truthfulness or falseness without a metaphysical base which allows us to support a realistic gnoseology? In other words, only an assessment of metaphysics as primary wisdom about reality can allow us to support the basic issues of knowledge. For example, the principle of non-contradiction is the basis of logical argument, and its critical examination regards metaphysical problems about reality properly; the value of knowledge about real things demands considering intentionality, that is, the apprehension of the way of being in an immaterial way; and, in the same line, realism is unsustainable unless the continuity between sensible and intellectual knowledge is admitted, which is given due to the abstraction process …” Artigas does not even accept that the criterion as proposed by Popper is enough to characterise experimental science in itself, because he says that Popper does not study how to build the scientific object. For Artigas, the experimental method depends, essentially, on this issue.

For Artigas, metaphysics has the last word on the criterion of demarcation. He says: “The examination of knowledge in itself demands placing oneself under a perspective which transcends such an approach (referring to experimental science, which he says is particular), because, in this case, what must be explained is how we can establish intentional contact with reality, which is the scope of the principles on which we rely, and which is the foundation of the method used. They are gnoseological questions that, in fact, are solved in their basic aspects by the realism bound to experimental science.”

For Artigas, the only perspective coherent with the radical weakening of metaphysics is the instrumentalist one. Artigas says: "Certainly, the demarcation exists, but it does not have the character proposed by Popper. Metaphysics and experimental science differ in their viewpoints … the building of the scientific object implies adopting particular perspectives realised in the type of definable properties of each objectivation. However, the metaphysical perspective leads to the total reality, without the limits implied by particular objectivations … For instance, the principle of causality expresses a general condition which is accomplished in reality, and therefore is also a condition of possibility of our knowledge: it refers to the necessity that all that happens is the result of causes. The particular causes, which can be of different types, are not specified; nor if they are necessary or free, determinist or not, either; it is only asserted that the natural results must be the result of causes. It is a modest assertion regarding details, totally general in its scope, and completely certain. It cannot be proved by the procedures of experimental science, but this fact does not prevent it from its validity; on the contrary, the validity of any knowledge, including that of science, is based on the principle of causality. It is not merely speculative either, because it is the result of the analysis of what experience shows us, though it transcends what is given in the experience because it leads to the rational explanation of reality."

Artigas finishes his comments about (4) by saying: "Science and metaphysics are joined from the beginning. That's why it is illusory to draw a line of demarcation that sets them in isolated spaces … the demarcation in itself supposes a metaphysical perspective."

As regards (5), Artigas claims that the unifying approach between science and metaphysics typical of realism has, as its main antagonist, scientificism. Quoting Radnitzky, he says that scientificism is characterised as being "the dogmatic belief that the way of knowing called 'science' is the unique which deserves the title of knowledge …" For Artigas, this is a description of an extreme scientificism, because he feels other softer type that would give science the title of the best knowledge among all the possible ones. This is Popper's approach, according to Artigas. When that type of knowledge is given priority in this way with respect to the other ones, Artigas states, it is the methodology that has the final word regarding the value of knowledge. In addition, it is a methodology alien to general metaphysical considerations with validity by themselves." As a corollary, Artigas says the following: "All knowledge has to be considered as conjectural, provisional, and refutable," and, therefore, "this solution ... is unable to explain the real value of scientific knowledge."

For Artigas, "scientificism is intrinsically incoherent, because it is not possible to assert that experimental science is the paradigm of valid knowledge if we cannot even prove the truthfulness of its statements. The only possible alternative consists in admitting a realistic metaphysics than can be the basis of a gnoseology in which the notion of truthfulness makes real sense."

In terms of (6), Artigas says that while science is socially considered to be the most reliable way of knowledge in existence, "the systematic study of philosophy of science seems to have placed thick shadows on the theoretical foundation of such a confidence". While Popper underlines that "the request for scientific objectivity makes it unavoidable that all scientific statements are provisional for ever …", Kuhn underlines that there are sociological explanations behind the agreement among scientific issues. Artigas responds in this way: "In both cases, the basic problem is that we do not manage a relationship between scientific statements and the realism of knowledge. Epistemology is reduced to methodology, and methodology to logic by Popper and to sociology by Kuhn."

Artigas says that “the term ‘reliability’ means the guarantees that, bearing in mind the achievement of a specific aim, something offers”. Applied to experimental science, it refers to the guarantees of methods and theoretical contents in order to reach the goal of scientific activity, and also the extent to which that activity fulfils certain purposes. According to Artigas, if scientificism is not overcome, it will not be possible to face the problem of reliability suitably because “the value of scientific knowledge depends on a philosophical base which transcends the experimental methodology and its foundation”. He concludes: “Experimental science does not equal to a cosmovision.”

Chapter 2. OTHER LEGAL AND SCIENTIFIC BASES

2.1 General epistemology foundations to prove facts

Most of the main ideas to introduce this topic can be found in (Lucena, 2016, pp. 141–145).

Uncertainty is a state of mind that can be reached after an intelligence operation called an ‘act of knowledge’. Knowledge theory is, undoubtedly, one of the subjects that every forensic expert should know and master – with the terminological background developed for ages and from discussions among scholars – to understand properly the following question they have to answer very often in trials: how confident are you about the statements made in the conclusions of the expert report? The court is interested in knowing the degree of confidence, confirmation or firmness of the judgement made by the expert in the conclusion of his report. All those ways of speaking express the same thing: the degree of uncertainty as a mental state in front of the result of an act of knowledge: the expressed assertion of a judgement.

General epistemology or knowledge criticism is a key element to go into the problem that we are trying to tackle with a solid rational argument: how to prove facts. And within that discipline the concept of certainty is classically dealt with.

Certainty is also a state of mind that may be achievable after the intelligence operation called ‘act of knowledge’. Unlike uncertainty, the firmness degree achievable in this case is the maximum. It does not make sense to explain a concept such as uncertainty in depth starting from something different to the explanation of certainty. In fact, uncertainty is essentially the lack of certainty. If the concept of certainty did not exist, if we did not know that it is a possible and determined state of mind, it would not be possible to define uncertainty, this word would not make any sense for us. Darkness cannot be defined without having previously defined brightness, or wickedness without goodness, or ugliness without beauty. The negative concepts stated as a lack of certain perfection are only intelligible as long as that perfection is previously defined and known.

Therefore, the definition of certainty in the context of proof of facts has an outstanding relevance. If certainty is defined as perfection achievable by the human mind when it is put in action, it has the quality of fullness. It is the direct effect on mind of something acting as cause: evidence. Therefore, uncertainty has essentially a limited value; it implies the fear of making a mistake (Millan, 2002, p. 116).

The state of mind in front of an act of knowledge has a psychological nature and the human being is fully aware of that: either it is satisfied or unsatisfied. The mind is satisfied when it believes it is right in asserting a judgement. If not, fear appears. Here there is a new and

significant concept for our purpose: how does the human mind know that it is right when asserting a judgement? It is the so-called 'truth criterion' or 'certainty criterion', depending on authors.

The 'truth criterion', which could be considered the master key to solving the problem from a rational viewpoint solidly, is evidence. Though several criteria have been suggested throughout history, none of them has overcome evidence as the ultimate criterion of certainty (Vernaux, 1971, pp. 147–150).

On the other hand, some outstanding philosophers of law have studied the epistemological model of judicial determination of facts according to the evolution of the predominant theories of knowledge in each historical stage (Lucena, 2016, p. 138).

The coherentist and pragmatist theories of truth were predominant in the second half of the 20th century Anglo-American epistemology (Moros, 2003, pp. 633–671) and its influence can be appreciated in the Bayesian literature.

The concept of belief has been used by Anglo-American epistemologists as an essential element in defining the concept of knowledge (see Lucena, 2016, pp. 138–140, which sums up the above-mentioned quote by Moros).

According to general legal doctrine about criteria for interpretation of legal texts, the grammatical meaning of 'legal texts' is the starting point to interpret any norm. However, it is also important to bear in mind the context in which those terms are used, which is known as the logical or systematic criterion. Both criteria, grammatical and logical, are mutually involved in our case. Therefore, legal terms used in laws of criminal proceedings or analogue regulations need to be understood by procedural actors respecting those criteria unequivocally.

In Chapter 4 all the above-mentioned epistemological concepts will be studied and discussed in greater depth.

2.2 Exhibits

At this point, it seems wise to define a terminology that will enable a better understanding and differentiation of the above-studied meanings and terms used to designate them.

Any standard needs a detailed and precise glossary of terms and, perhaps, the first thing to bear in mind before doing it is the existence of a procedural context (in this case focused on the criminal jurisdiction) conditioning, in some way, the terminology that we could consider to be suitable for our purposes. Expert reports have the legal nature of a documented legal act submitted in a criminal proceeding. It is a technical-scientific act and, therefore, the consequences of such a documented act are answers to questions made by petitioners. Those answers include the results of the analyses and all kinds of interpretations derived from them.

Well, if the aim of our study is to standardise the terminology to be used in conclusions of forensic reports, it is thought necessary to start with one most directly related to scientific tests. And the following basic terms appear in its development: collected material in the crime scene sent to the laboratory to be analysed, control or reference material, analytical procedure applied to materials and the results of such a procedure.

Collected material can be called, in Spanish, either *huella* (print), *vestigio* (vestige), *indicio* (circumstantial evidence), *rastro* (trace), *señal* (sign) or *traza* (track, trail) when they are consequences of the action of one material on another, or *muestra* (sample), *vestigio* (circumstantial evidence), *indicio* (circumstantial evidence), *resto* (remains), *objeto* (object), *efecto* (effect) or *artículo* (item) when they are properly collected material. However, there is a more precise legal term known as *pieza de convicción* (exhibit) (Clemente, 2000). Though this term is also applied to material seized by the police, material collected after crime scene investigation or police investigations and submitted in the proceedings are designated using the same term.

Control or reference material is also called 'collected material' if it is mentioned in a general way. If the material is useful for making a personal identification it is commonly called *reseña* (police record), with the adjective 'undoubted' often being added. Reference material that identifies an element of the periodic table or a compound is called *patrón* (pattern).

Laboratories work with materials submitted by investigation units and judicial authorities. The multiplicity of terms in each language and judicial system to describe them and the need to emphasise some of their properties make it advisable, in our view, to choose a higher level than the observational one to designate them. The level that is suggested for designating materials is the specific one of legal science: all those materials, regardless of the properties to be highlighted, are 'exhibits'. In our view, a standard on conclusions of expertise has to do with information directly examined by judges in the proceedings (a conclusion is the part of the expert report that they always read). If so, since materials submitted to laboratories belong to criminal proceedings, it is convenient to name such materials as clients of expert reports name them in the proceedings. The English word 'exhibit' has the same meaning as the Spanish expression *pieza de convicción*.

In the case of using adjectives to describe an action on a material or a property of it (i.e. whether it is collected, or is a reference, or is compared to something relevant, observed, examined or analysed, or is coming from a known or unknown source, and so on) it is believed that there will not be any loss of intelligibility if the word 'exhibit' is used (presenting in its legal meaning the very generic sense of 'thing' or 'material'), attaching to it the relevant qualifications.

2.3 Statistical evidence

The word 'evidence' is extremely polysemic in English. One of its meanings is "something such as a fact, sign or object that gives proof or reasons to believe or agree with something" (Longman dictionary, ed. 1987). According to this definition, our "fact, sign or object" is scientific observations (findings) and those things that observations refer to are a pair of mutually exclusive propositions.

From the scientific viewpoint, the statistical meaning of 'evidence' according to the so-called law of likelihood was defined as follows (Royall, 1997, p. 3):

"If hypothesis A implies that the probability that a random variable X takes the value x is $p_A(x)$, while hypothesis B implies that the probability is $p_B(x)$, then the observation X = x is evidence

supporting A over B if and only if $p_A(x) > p_B(x)$, and the likelihood ratio, $p_A(x) / p_B(x)$, measures the strength of that evidence" (Hacking, 1965)[3].

The following question expresses, in Royall's opinion, the context where we should interpret the meaning of 'evidence':

"What does this observation tell me about A versus B? (How should I interpret this observation as evidence regarding A versus B?)"

Royall says, in the paragraph titled 'Relativity of evidence' on page 8, the following: "The law of likelihood applies to pairs of hypothesis, telling when a given set of observations is evidence for one versus the other: hypothesis A is better supported than B if A implies a greater probability for the observations than B does. This law represents a concept of evidence that is essentially relative, one that does not apply to a single hypothesis, taken alone."

Royall wonders whether a valid rule that will guide the interpretation of statistical data as evidence relating to a single hypothesis can be found, without reference to an alternative. He examines two candidates: he calls the first 'the law of improbability'. It states that X = x is evidence against A if $p_A(x)$ is small, that is, if A implies that the observation is improbable. He calls the second 'the law of changing probability', and it states that X = x is evidence for or against A according to whether the effect of the observation is to increase or reduce the probability that A is true. He argues that neither of these rules represents a satisfactory concept of evidence for scientific discourse, the first because it is wrong (after examining the rationale for testing significance), and the second because it is subjective (although stated in terms of a single hypothesis, and not referring to any explicit alternative, it actually entails both alternative hypotheses and conditions on how prior probability is distributed among the hypotheses).

There are several important reasons for accepting the law of likelihood as a basic framework to define the concept of evidence scientifically (section titled 'Towards verification', pp. 5–7):

a) "It seems to be the natural extension, to probabilistic phenomena, of scientists' established form of reasoning in deterministic situations: if A implies that under specified conditions x will be observed, while B implies that under the same conditions something else, not x, will be observed, and if those conditions are created and x is seen, then this observation is evidence supporting A versus B. This is the law of likelihood in the extreme case of $p_A(x) = 1$ and $p_B(x) = 0$. The law simply extends this way of reasoning to say that if x is more probable under hypothesis A than under B, then the occurrence of x is evidence supporting A over B, and the strength of evidence is determined by how much greater the probability is under A." And Royall comments: "This seems both objective and fair – the hypothesis that assigned the greater probability to the observation did the better job of predicting what actually happened, so it is better supported by that observation."

b) The law of likelihood is consistent with the rules of probability theory. After X = x is observed, the a priori probability ratio Pr(A) / Pr(B) is changed following the odds version of Bayes' Theorem to the a posteriori ratio Pr(A| X = x) / Pr(B| X = x). This shows that the new evidence, X = x, changes the probability ratio by the factor $p_A(x) / p_B(x)$, precisely in agreement with the law of likelihood. "If we use the law then our interpretation of data as evidence will be consistent with the rules of the probability theory: we will never claim

[3] $p_A(x)$ and $p_B(x)$ are conditional probability distributions.

that an observation is evidence supporting A over B when the effect of that observation, if A and B had probabilities, would be to reduce the probability of A relative to that of B."

c) "If we use the law of likelihood, will we be led to the truth? Suppose A is actually false and B is true. Can we obtain observations that, according to the law, are evidence for A over B? Certainly. Does this mean that the law is invalid? Certainly not. Evidence, properly interpreted, can be misleading. This must be the case, for otherwise we would be able to determine the truth (with perfect certainty) from any scrap of evidence that it is not utterly ambiguous. It is too much to hope that evidence cannot be misleading. However, we might reasonably expect that strong evidence cannot be misleading very often. We might also expect that, as evidence accumulates, it will tend to favour a true hypothesis over a false one more and more strongly. These expectations are met by the concept of evidence embodied in the law of likelihood." Royall proves mathematically on pages 8 and 9 the fulfilment of those expectations, and he quotes this reference: Robbins, 1970, pp. 1397–1409. See APPENDIX II on pages 157–158 for further details.

In Section 1.11 titled 'Irrelevance of the sample space', Royall says that the law of likelihood states that the evidence in an observation, X = x, as it pertains to two probability distributions labelled θ_1 and θ_2 is represented by the likelihood ratio, $f(x; \theta_1) / f(x; \theta_2)$[4]. In particular, the law implies that to interpret the observation as evidence for hypothesis H_1: $\theta = \theta_1$ vis-à-vis H_2: $\theta = \theta_2$, only the likelihood ratio is relevant. What other values of X might have been observed and how the two distributions in question spread their remaining probability over the unobserved values is irrelevant – all that counts is the ratio of the probabilities of the observation under the two hypotheses.

Another important concept related to evidence is the so-called 'likelihood principle'. This principle asserts that two observations that generate identical likelihood functions are equivalent as evidence. Royall quotes the following words from Birnbaum: "the 'evidential meaning' of experimental results is characterised fully by the likelihood function" (Birnbaum, 1962, pp. 269–306).

As an example to understand this principle, Royall explains in Section 1.12, titled 'The likelihood principle', the following: "Suppose two simple hypotheses for the distribution of a random variable X assign respective probabilities $f_1(x)$ and $f_2(x)$ to the outcome X = x, while two different hypotheses for the distribution of another random variable Y assign respective probabilities $g_1(y)$ and $g_2(y)$ to the outcome Y = y. If $f_1(x) / f_2(x) = g_1(y) / g_2(y)$ then the evidence in the observation X = x regarding f_1 vis-à-vis f_2 is equivalent to that in Y = y regarding g_1 vis-à-vis g_2." This process may be repeated at our convenience using more outcomes and distributions and is called the likelihood principle. It is usually stated in terms of likelihood functions.

The law of likelihood gives the likelihood function[5] its meaning: if $L(\theta_1; x) > L(\theta_2; x)$, then the observation x is evidence supporting the hypothesis that θ is θ_1 (that is, the hypothesis that X has the distribution identified with the parameter value θ_1) over the hypothesis that θ is θ_2 (that is, the hypothesis that X has the distribution identified with the parameter value θ_2), and the likelihood ratio $L(\theta_1; x) / L(\theta_2; x) \equiv f(x; \theta_1) / f(x; \theta_2)$ measures the strength of that evidence.

[4] $f(x; \theta_1) / f(x; \theta_2)$ can be written as $f(x \mid \theta_1) / f(x \mid \theta_2)$. They represent probability density functions.

[5] For a fixed value x, $f(x; \theta)$ can be viewed as a function of the variable θ and it is then called the 'likelihood function'. We will use the notation $L(\theta; x)$ for the likelihood function.

In essence, Royall suggests that the concept of statistical evidence is properly expressed in the law of likelihood and that the likelihood function is the appropriate mathematical representation of statistical evidence (extracted from the section titled 'Evidence and uncertainty', p. 28):

"Probabilities measure uncertainty and likelihood ratios measure evidence. A probability density function represents the uncertainty about the value of a random variable; it describes how the uncertainty is distributed over the possible values of the variable (the sample space). That uncertainty disappears when the observation is made – then the value of the random variable is known, and that value is evidence about the probability distribution. The likelihood function represents this evidence; it describes the support ratio for any pair of distributions in the probability model."

And this is the summary (Section 1.14, p. 31):

"The question that is at the heart of statistical inference – When is a given set of observations evidence supporting one hypothesised probability distribution over another? – is answered by the law of likelihood. This law effectively defines the concept of statistical evidence as being relative, that is, a concept that applies to one distribution only in comparison to another. It measures the evidence with respect to a pair of distributions by their likelihood ratio. The law of likelihood is intuitively reasonable, consistent with the rules of probability theory, and empirically meaningful."

2.4 Forensic evaluation of evidence: the Bayesian perspective

This thesis is based on a normative viewpoint about the logic of uncertainty reasoning applied to forensic evaluation of evidence (see Section 4.2.1 for further details). In addition, according to the Bayesian subjectivism, this author holds that the evaluation of scientific evidence in court is the expression of one's personal degree of conviction (Taroni et al., 2006, p. 20).

Modern Bayesianism comes from the papers written by F.P. Ramsey (1931, pp. 156–198) and B. de Finetti (1930, pp. 258–261) in the thirties of the last century. L.J. Savage spread and promoted the new approach from the fifties (1954). The contributions of C.R. Kingston and P.L. Kirk (1964, pp. 514–521), and J.B. Parker and A. Holford (1968, pp. 237–251), in forensic science in the sixties were relevant. D.V. Lindley (1965; 2007) stands out in the middle of the sixties and the seventies though his writings have influenced forensic science thus far. Since then, many other relevant authors have appeared, but in terms of the topics tackled in this thesis, the contributions of R.C. Jeffrey (1983), J. Pearl (1982, pp. 133–136), S.L. Lauritzen and D.J. Spiegelhalter (1988, pp. 157–224) should be qualified as essential in developing the Bayesian networks from the eighties. Ian Evett was awarded with the first Distinguished Forensic Scientist Award in 2000, as the leading author of a series of outstanding publications in the field of forensic science for the period 1997–1999 on a model of case assessment and interpretation, published in *Sciences and Justice* in 1998. The contribution of the co-authors R. Cook, G. Jackson, P.J. Jones and J.A. Lambert was also appreciated. B. Robertson and G.A. Vignaux wrote the book *Interpreting Evidence – Evaluating Forensic Science in the Courtroom* (1995), and C.G.G. Aitken and F. Taroni (1995) published the first version of their book *Statistics and Evaluation of Evidence for Forensic Scientists* in the nineties, which lots of scholars consider to be masterpieces in forensic evidence evaluation.

Efficient proposals to compute numerical likelihood ratio values in forensic comparisons have been published since the eighties of the last century. Nevertheless, the vast majority of their practical forensic applications have been spread worldwide in the current century, for example in the following forensic fields: DNA analysis (B.S. Weir, 1996; I.W. Evett et al., 1998; J.S. Buckleton et al., 2004; D.J. Balding, 2005; J.M. Butler, 2005); speaker recognition (D. Meuwly, 2001, PhD; J. Gonzalez et al., 2006, pp. 331–355; 2007, pp. 2072–2084; D. Ramos, 2007, PhD); comparisons of fibres, paint fragments or glasses (I.W. Evett, 1984, pp. 25–32; S.D. McDermott et al., 1997, pp. 1012–1018; 1999, pp. 263–269; C. Champod et al., 1997, pp. 75–83; J.M. Curran et al., 2000; J. Zieba et al., 2008, pp. 47–58; G. Zadora et al., 2014); comparisons of shoeprints, footprints or fingerprints (C. Champod , 1996, PhD; I.W. Evett et al., 1998, pp. 241–247; F. Taroni et al., 2002, pp. 15–25; R.B. Kennedy, 2003, pp. 53–63; C. Neumann et al., 2007, pp. 54–64); toolmarks and firearms (S.G. Bunch, 2000, pp. 955–962; S.G. Bunch et al., 2013, pp. 223–229); isotopic analysis of explosives (G. Pierrini et al., 2007, pp. 43–48); combustion accelerator analysis (G. Zadora et al., 2014; P. Vergeer et al., 2014, pp. 401–411); gunshot residue (GSR) analysis (A. Biedermann et al., 2009, pp. 24–35; 2010 pp. 103–110), and others.

The Bayesian approach is a solid logical framework to make probabilistic inferences, that is, it can be used even if numerical data is not available. In these cases it would be advisable to follow the recommendations made by O'Hagan et al. in their work titled *Uncertain Judgements: Eliciting Experts' Probabilities* (2006). In this sense, it is worth quoting what G. Shafer once said (Taroni et al., 2006, p. 2): "Probability is not really about numbers; it is about the structure of reasoning", and uses qualitative assignments.

As a forensic practitioner applying an automatic speaker recognition system to forensic audio recordings, the author is indebted to an experimental work carried out in 2005 by J.B. Tapias (Tapias, 2005) after the first three official forensic reports done in the Acoustics Area of the Criminalistics Service of the Spanish Civil Guard using the first version of BATVOX (www.agnitio-corp.com). His experimental results using the speech databases available in this unit were important to the personal understanding of this author of what N. Brummer and J. Du Preez (Brummer et al., 2006, pp. 230–275) and D. Ramos (2007, PhD) wrote about calibration of LRs afterwards. The huge variability of the samples of speech in speaker recognition using automatic systems advised researchers to maximise the evaluation of the accuracy of the LR values in the NIST Evaluations organised by the Department of Commerce of the United States from 2004 on. This explains the original idea of this author to explore the use of Tippett plots to detect uncalibrated LRs in accordance with the results obtained by J.B. Tapias (2005).

2.5 Automatic speaker recognition systems: forensic applications

The Acoustics Area of the Criminalistics Service of the Spanish Civil Guard was one of the first official forensic laboratories in the world to use an automatic speaker recognition system to compare voices. The first version, called IdentiVox, was developed by ATVS in 1998. First, this version was used in the context of a close-set comparison related to wiretaps since December 1999. Assuming that the information obtained by the police investigating the crime deserves credit, a final decision on such an expectation was made according to the logical framework used (Neyman-Pearson significance tests). The requisites for such a decision were as follows: a wiretapping requires justification before a judge as to why some telephone numbers need to be tapped to investigate a crime; the main speakers around the wiretapping are usually under surveillance for a long time; and forensic practitioners check whether by training a GMM (Gaussian Mixture Model) model per speaker (suspect) using speech from different

conversations assigned to him/her by the transcriber, by means of a close-set classification test and a later t-Student test to compare the classification error rate of the proof with the expected one, previously obtained using ad hoc experimental databases, the results are comparable or not.

The above-mentioned procedure changed completely once the Bayesian approach was selected as the right one and was implemented in a new version of IdentiVox called IdentiVox-LR in 2004. Three years were needed to develop this implementation because the sources of variability involved in speaker recognition comparisons are huge. ATVS was inspired by D. Meuwly's work (Meuwly, 2001, PhD) to implement the likelihood ratio approach in their systems, and has participated in the NIST Evaluations of the Department of Commerce of the United States since 2000.

The first tested and effective forensic version of IdentiVox-LR had the following main technical characteristics:

a) Text-independent acoustic-spectral short-term system based on MFCC – Mel-Frequency Cepstral Coeffecients (Davis et al., 1980, pp. 357–366; Plumpe et al., 1999, pp. 569–586);
b) GMM-UBM (Gaussian Mixture Models – Universal Background Model) generative modelling approach (Reynolds et al., 2000, pp. 19–41);
c) Maximum a posteriori adaptation – MAP of the speaker GMM from the UBM;
d) Z-normalized scoring to have a 'common' non-targeted distribution for all speakers;
e) T-normalized scoring (Auckenthaler et al., 2000, pp. 42–54) as a kind of test duration normalization.
f) CMN-RASTA-Warping channel-normalized scoring techniques to compensate for the huge influence of channel variability (Furui, 1981, pp. 254–272; Hermansky et al., 1994, pp. 578–589; Pelecanos et al., 2001, pp. 213–218).

ATVS technology was transferred to AGNITIO S.L. (www.agnitio-corp.com) in 2004. The first technological core for speaker recognition was called COREVOX 1.0.

After detecting some problems in solving the first three real cases (see Section 7.5), D-normalized scoring (Ben et al., 2002, 689–692) was additionally implemented in the first commercial version of IdentiVox-LR called BATVOX to avoid the dependency of scores on model training lengths.

From 2002 to 2006, major advances in the field of pattern recognition technique were achieved:

- Support Vector Machines – SVM (Schölkopf et al., 2002)
- Nuisance Attribute Projection – NAP channel compensation (Solomonoff et al., 2005, pp. 629–632)
- Joint Factor Analysis – JFA channel compensation (Kenny, 2005)
- Gaussian Supervector – GSV-SVM (Campbell et al., 2006, pp. 308–311)

Though this new technology improved the discrimination power of the automatic speaker recognition systems, only NAP channel compensation was implemented in the third version of BATVOX.

The papers of ATVS (Gonzalez et al., 2006; 2007) were significant both in the international specialised forensic community and in the main international conferences and journals promoted by IEEE (Institute of Electrical and Electronics Engineers).

The Acoustics Area of the Criminalistics Service of the Civil Guard achieved the accreditation of forensic voice comparisons under the ISO 17.025 norm in 2013. It was the first European Forensic Area to obtain it. The AGNITIO product called BATVOX 3.0 was used.

From 2006 on, new relevant advances were obtained:

- i-vector front-end extraction (Dehak et al., 2011, pp. 788–798)
- Probabilistic Linear Discrimination Analysis – PLDA modelling and scoring (Prince et al., 2007, pp. 1–8)
- I-vector length normalization (Garcia Romero et al., 2011, pp. 249–252)

Though BATVOX 4.0 includes i-vector-PLDA technology, its forensic use has not been put into effect by forensic practitioners belonging to the Criminalistics Service. This is the current challenge now that the corresponding training courses are finished.

Nonetheless, this overview of the technological evolution of an automatic speaker recognition system would be incomplete without quoting the advances in measuring their performance in computing likelihood ratios. In this sense, some outstanding scientific contributions should be highlighted, such as the paper first published in the Proceedings of Odyssey '04 – the Speaker and Language Recognition Workshop held in Toledo (Spain) by N. Brümmer titled 'Application-Independent Evaluation of Speaker Detection', and later on extended and improved (Brümmer et al., 2006), and some papers of D. Ramos about the topic of measuring calibration of likelihood ratios (Ramos et al., 2013a, pp. 156–159; 2013b).

PART II: PRACTICAL ASPECTS

Chapter 3. THESIS

3.1 Proposition 1

The logic of concepts based on Aristotle-Thomas's metaphysics prevents us from falling into nominalism: the universal predication of concepts is based on the real composition of things. The inherent realism in that school of thought is considered essential to ensure that concepts used in the conclusion of forensic reports are intelligible by courts and juries, and coherent with the actual use given to those concepts in the judicial scope.

Nominalists consider ideas to be mental fictions, useful for classifying objects, but useless for reaching true knowledge. They are mere common names to call things, and they lack intelligible content. It is only possible to know if things exist or not (intuitive knowledge), but beyond this point will be more or less plausible explanations to make any sense of reality. There is explicit renunciation to know what things are, and special attention is paid to discovering natural laws.

According to this way of thinking, instead of feigning explanations of reality, philosophy had to change to find a new and solid base to be a proper knowledge. Aristotle's metaphysics of knowledge stated that evidence is the cause of certainty in the mind, and evidence is a property of the object. Descartes, however, equalled evidence and certainty. He set a prerequisite to check which ideas deserve confidence and which do not. Descartes lost the intentionality of knowledge, focusing his attention on the object, and questioned whether it represents something real. He replaced intentionality with causality. Ideas are caused, and they refer to reality as a cause, not as an object.

To know is not to open oneself to reality, to accept reality as it is, but to control reality. Before accepting any object, it has to be exhaustively and previously examined to be accepted as clear and distinct by the mind.

Empiricism is nominalism, and Kant thought that to know is, in short, to build the object of knowledge because he was influenced by rationalism as well. After Kant, many philosophers thought that realism was not possible, and Hegel maximised the criticism.

Scientists have always desired to be as objective as possible and many found the most credible philosophical reasoning to their beliefs in empiricism at the beginning of the last century. All true knowledge should be based on empirical verification, they said. Therefore, the logic positivism formulated the proposal that a theory or law should be verified to be accepted as true. The end of this way was that they had to accept a completely subjective and utilitarian concept of truth, but the judicial process to fix facts is not compatible with such a concept of truth, but with a concept of truth based on existence.

3.2 Proposition 2

A test may be qualitative, quantitative or comparative. The result of a qualitative test is a proposition (including false positive and negative error rates), the result of a quantitative test is a numeric figure (including an associated uncertainty measurement), and the result of a

comparative test is a likelihood ratio (including the performance evaluation of the system when it is numeric).

Results are interpreted by means of a procedure legally known as expert or scientific evidence, which is included within the inductive or indirect evidence according to its epistemological nature. In this context, an expert interprets results as either merely technical or evaluative interpretations. In the latter, the expert interprets data as evidence according to the propositions supported by the parties, and also to what he has a priori assumed regarding the nature of the problem to be solved and derived from his knowledge and experience as expert, and what he a priori knows about the circumstances of the case. The interpretation of data as evidence just described and carried out by an expert has been called the 'law of likelihood' (Hacking, 1965). This way of evaluating results of tests by means of likelihood ratios has been called the 'likelihood paradigm' (Royall, 1997), in contrast to the individualisation paradigm in which the evaluative character looked for interpreting data as evidence according to the legal principle of equality of arms does not fit.

In our opinion, the English term 'evidence', used as a neologism in Spanish (literally translated as *evidencia*), would add still more confusion to the current difficult interpretative problem related to how experts have to communicate the results of their reports to the court. Its highly polysemic level in English where, for example, the term 'evidence' is applied either to the collected material after a visual inspection of the scene of the crime, or to the quantitative or comparative results obtained once the appropriate analyses have been carried out, or even trying to express the evaluative character of an interpretation of the results of any test we say that data are evaluated as evidence, so all the above-mentioned reasons make it difficult for the court to understand what the expert tries to communicate in his report. In our view, it would be better to distinguish among exhibits, test results, interpretation of results, and within the latter between merely technical and evaluative ones.

The most predominant official forensic practice in the world is still linked to the paradigm of individualisation. This paradigm has been inserted in the judicial practice by forensic experts belonging to official laboratories. The shift in paradigm, called the 'likelihood' paradigm, promoted from the academic scientific scope, has to be produced mainly in the same way because of the very nature of the judicial system.

Few official forensic laboratories in the world have a staff mainly constituted by scientists and usually forensic practitioners carry out their tasks according to protocols accepted by the international forensic community at best. Jurisprudence is the common tool for settling disputes related to evidence in the Continental law. In contrast, there are rules of evidence in the Anglo-Saxon law. It is surprising that in this topic, both systems differ in the opposite direction to most topics.

The so-called 'identification' paradigm may be characterised by the belief that the forensic practitioner is the person who should solve the problem of knowing whether two compared samples, one taken from the suspect (undoubted) and the other from the scene of the crime (doubted), have the same origin, or not. That is the very reason that justifies the role of experts assisting courts to measure the evidence value of scientific proofs, according to most opinions in legal and law enforcement scopes.

The Bayesian approach applied to the evaluation of evidence in courts rejects such a belief. According to the theory of probability, such belief lacks a scientific base. The 'identification'

paradigm can be described as "… a faulty probabilistic intuition equating infrequency with uniqueness" (Saks et al., 2008, pp. 199–219). In addition, as expressed in (Lucena et al., 2012, p. 71)[6], "It has been a key determinant to achieve convictions, as prosecutors and law enforcement agencies have relied on reporting results including categorical statements about identity on which to base their theses."

Though the Spanish judicial system clearly establishes that the main role of the Prosecutor in penal proceedings is to exercise the legal action on behalf of the State in every crime (Art. 124.1 of the Spanish Constitution of 1978; Art. 3.4 of the Statute of the Attorney General's Office; Art. 105 of the Law of Penal Proceedings), representing the general interest to carry out justice, it is not obliged to hold the charge if there is no basis for doing so, and it has to bear in mind all the concurrent circumstances in the penal fact, in favour or not of the defendant (Art. 2 of the Law of Penal Proceedings). The Prosecutor can ask the investigative judge for police inquiries to verify the commission of a crime, and to investigate the responsibility for the crime (Art. 3.5 of the Statute of the Attorney General's Office).

On the one hand, the above-mentioned legal jurisdictions incline Prosecutors to ask official forensic laboratories to confirm their hypotheses, mostly against the accused. On the other hand, law enforcement agencies try to clear up crimes by arresting criminals. The well-known and expected result is that most forensic examinations are requested by the prosecution. For this very reason, the 'likelihood' paradigm is an excellent antidote to counteract the quoted tendency. This paradigm avoids considering official forensic laboratories as prosecution laboratories, even if they belong to police bodies.

The personal experience of the author of this doctoral thesis, who has been working in the Central Laboratory of the Spanish Civil Guard in Madrid (Spain) for more than 26 years, is that the police mentality always prevails over the scientific mentality in the staff – mainly if those who are in charge of the different internal organisation levels are not scientists. Therefore, it should be emphasised that forensic institutes, with staffs not mainly made up of scientists, will have to adopt special measures to counteract the accusative tendency in the minds of their personnel.

3.3 Proposition 3

The application of the 'likelihood' paradigm by courts requires specific statistical technical assistance from the experts. The inferential reasoning to assess the proof of facts judicially needs to become more technical to ensure its logic quality. This makes up an important additional epistemological guarantee in this evaluative process since it is considered to imply a more efficient application of the principle of the equality of arms in the judicial process.

The 'likelihood' paradigm in forensic science separates perfectly the role of the court or jury from that of forensic experts. The odds version of Bayes' Theorem illustrates this by considering odds to be the role of the court or jury, and the likelihood ratio to be the role of the forensic expert. But this distinction cannot be understood as if forensic experts would not be allowed to help judges and juries to assign odds.

Gascon says the following with respect to the judicial decision of legal conflicts: "It is usually represented as a 'practical syllogism' wherein starting from a major premise (the suitable

[6] The page numbers of this reference will be written according to the pagination of this document because the paper has been included in Chapter 6.

judicial rule) and, at least, a non-mandatory minor premise (factual), a normative conclusion is derived." She continues: "The minor premise has to be a particular case of the more general case of the major premise. This factual premise is not a simple descriptive statement of an event. It is the result of a judicial activity by which some facts are qualified as a concrete case of the abstract case in which they have to be subsumed, and this activity of judicial qualification of facts has a normative nature" (Gascon, 2010, pp. 45–46).

But as Guastini says: "From a merely logic viewpoint it should be carefully distinguished between problems of knowledge of facts and problems of judicial qualification of facts. The former are empirical, and the latter are problems of interpretation" (Guastini, 1996, p. 201, fn. 30). This implies that the knowledge of facts should be tackled by applying the logical framework of the general epistemology. Therefore, the proof of facts, in its entirety, can be considered a field of scientific expertise.

Experts can help judges and juries to assign odds. The probabilistic inference involved in the assessment of data as evidence entails a closer collaboration between members of courts and juries, and experts. Assigning prior and posterior odds in the inferential process to measure the evidence value of a scientific proof is not a trivial matter.

Chapter 4. RESULTS AND DISCUSSIONS

The results take two different forms: Sections 4.1 and 4.2 are almost self-contained, though part of the former has ended up being the last published research of this thesis (see Chapter 9), and Sections 4.3, 4.4 and 4.5 summarise the results of previous research papers, all of them presenting practical applications of the propositions written in Chapter 6. Specifically, Section 4.1 addresses Proposition 1; Sections 4.2, 4.3 and 4.5 address Proposition 2; and Section 4.4 addresses Proposition 3.

4.1 Epistemology applied to conclusions of expert reports

The Abstract of the paper, published in *Forensic Science International* with the title of this section (see chapter 9), introduces this topic.

The aim of this study is only to illustrate this problem by comparing similar legal terms used in English and Spanish, and taking into account the Anglo-Saxon and Continental legal systems.

4.1.1 Metonymies

4.1.1.1 Doubt and uncertainty

One of the most basic and universal principles of the criminal proceedings law is expressed by the following Latin aphorism: *in dubio pro reo*. Everybody knows that it is related to mandatory decisions made by a court when pronouncing a verdict of innocent or guilty of a crime once all the mandatory stages of a criminal proceeding are finished. This doubt is part of a decision-making process.

However, doubt also exists when the mind is not determined regarding the truth of a judgement. The same term applied to both contexts (assent and decision) is justified because the same state of mind is trying to be described and it is characterised by a lack of determination to assert a proposition or to make a decision after the proposition has been asserted. This sort of indeterminacy implies the absence of an obligatory rational assertion or decision. Therefore, it also implies a situation of maximum uncertainty. If several options are available, all of them must be equally probable to be asserted or decided.

When doubt is related to making a judgement, it is said that the judgement is suspended (and the assent as well). When doubt is related to taking a decision, it should be analogously understood that there are no arguments for the court to be forced to consider the defendant more likely to be guilty than innocent.

However, the aphorism *in dubio pro reo* is understood by a lot of people as meaning that in a case where the court 'has some doubts' (it isn't completely sure) about a verdict of guilty, a verdict of innocent must be given. The inverted commas statement seems to be interpreted as if an adjustment on doubt could be admissible: 'to have some doubts', adjustment that is also noticed when people use another well-known interpretative legal principle: 'beyond all reasonable doubt'. The terms 'doubt' and 'uncertainty' become confused: to be in doubt is the same as being uncertain. However, while adjustments are compatible with uncertainty, is it the same with doubts? Not at all if we consider the etymological concept of doubt (whose Latin root is the verb *dubitare*) as a reference derived from the philosophical sceptical doubt of Pirron, referring to the state of mind of indetermination, and this is only possible within a

determined state of uncertainty: the equal probability of all possible options. In such a defined state of mind called 'doubt', uncertainty associated with a set context in which a number of known and finite options are determined is not adjustable.

Note the inappropriate transfer of meaning from 'uncertainty' to 'doubt'. The Latin aphorism *in dubio pro reo* is true based on the presumption of the innocent principle perspective, where 'doubt' is understood from its etymological (sceptical) meaning: in the case of indetermination, the court must pronounce a sentence of innocent of a crime.

When, instead of doubt, there is uncertainty, the court must make, and makes, in fact, the assessment by means of probabilities and must apply the so-called 'standards of proof' according to experts in the field. In that case, the aphorism should be understood with a meaning of 'doubt' beyond its etymology and therefore it is clearly a case of metonymy. It could be interpreted as if the court will do the same thing as when in doubt (giving the verdict of innocent) following the criteria established in the standard of proof regarding the overcoming of the reasonable uncertainty degree in a particular case to give the verdict of guilty.

'Doubt' is often used in Spanish as a word suitable for adjustment: for example, when we quote the saying 'I have absolutely no doubt' we express our certainty that the judgement is true, though in other circumstances a respectful interpretation would be possible with the etymological meaning of doubt: 'I have a little bit of doubt', where the focus is on the matter in itself: 'it is not highly relevant for the person speaking'. Therefore, 'the degree of relevance of the subject' is established for the decision in suspension.

There is a figure of speech called 'metonymy', which is more frequently used than people usually believe and which refers to the naming of something using another name somehow related to the former. When that relationship consists in taking a part of something as being referred to the whole thing or vice versa, it is called a 'synecdoche'. For example, nationalists usually designate themselves by the name of the inhabitants of the country or region where they live and who think like them.

This synecdoche about the word 'doubt' – a term with two different meanings: its own meaning, which is a part of something, and the related and attributed one, which is the whole thing, 'uncertainty' – is a linguistic problem in the forensic practitioner-court communication that can make the understanding of terms used in forensic report conclusions difficult.

'Doubt' and 'uncertainty' are synonyms in English. Therefore it is possible to measure doubt. This is an important semantic difference with the Spanish formal meaning of 'doubt'.

4.1.1.2 Certainty and uncertainty

A brief introduction to this subject is given in (Lucena, 2016, pp. 145–146).

Metaphysical certainty is absolute, as it derives from those laws of being that are necessary and cannot be overridden. As an example we can think of the certainty generated in our mind by the so-called first principles: non-contradiction and identity. This type of certainty includes the one achievable by mathematical science and even the certainty related to the intuition of simple facts as being conscious of our own existence (we have to exist if we are conscious of existing), or when a certain event takes place in our presence and we are later able to clearly

describe it (in such conditions, it is not possible that the event has never happened). In these later cases, intuition reveals a need in fact similar to the need of principles.

Certainty based on sensible experience is metaphysical certainty, rather than physical. The latter is based on a natural law while the former relies upon an ontological law. Due to the fact that knowledge of natural laws is inductive, the physical certainty limits are determined by the nature of this kind of reasoning, which is usually applied by experimental science and therefore is a kind of certainty that cannot be strictly considered more than opinion.

Moral certainty is based on a moral law. It comprises certainties based on both sociological and psychological laws as well. These laws are not as necessary as natural laws because we cannot forget that human beings are free. It seems wise to consider that this kind of certainty can never go further than opinion.

It is possible to reach certainty in those cases, and it is even valid to speak about a hierarchy in the strength of the truth of a judgement in mind depending on the type of certainty: firstly metaphysical, then physical and lastly moral. This type of adjustment in the strength of certainty is also called 'epistemological relevance' or 'quality of certainty'.

When we speak about adjustment on certainty we do not use the distinction mentioned above, which derives from scholars, but a metonymy is present again. In this case the term 'certainty' is used instead of 'uncertainty', the latter capable of adjustment, to designate the state of mind on the truth of a judgement in any situation. What is determined, the state of mind under certainty, characterised by the absence of fear to err, is subsumed in the concept of uncertainty, essentially characterised by the opposite. Again as a part of something, certainty, is taken for the whole, certainty plus uncertainty, therefore there is again a synecdoche.

The term 'certainty' has the same meaning in English and Spanish. It means the highest strength of the assent carried out by the mind on the truth of a judgement or when making a decision. It is wrong to speak about adjustment on certainty in both languages.

As Lindley stated (Lindley, 2014), although in practice the phenomenon of uncertainty is a real fact familiar to any single human being, general epistemology points out that the notion of certainty is not always well understood. Intelligence can consent in two different ways: (i) when it is caused by the known object; and (ii) when it is biased to a certain judgement by will. In the first case we can find two situations: (i) when the object is known immediately – what we call 'intelligence'; and (ii) when the object is known through something else, such as, for instance, a demonstration – what we call 'science' (there are two types of demonstration: *quia* or inductive (from effects to causes) and *propter quid* or deductive (from the causes or nature of something to effects or properties)). In the second case we can also find two different situations: (i) when the judgement is produced along with some reservations – then we call it an opinion and therefore uncertainty is present in our mind; and (ii) when the judgement is expressed in absolute terms – what we call an act of faith (Vernaux, 1971, pp. 137–138).

4.1.1.2.1 Opposition type between certainty and uncertainty

Aristotle defined opposed concepts as "those which mean attributes which cannot inhere in the same subject at the same time" (Aristotle, Metaphysics). And he distinguished four opposition types: contradictory, privative, contrary and relative.

Since intelligence tends to naturally know reality it tends to achieve the state of certainty. Uncertainty is, in this context, a deprival of an owed characteristic to intelligence. Therefore, the most appropriate opposition in relation to the pair of concepts (certainty/uncertainty) of our interest is the so-called 'privative'.

Further details on this topic can be found in (Lucena, 2016, p. 146).

4.1.2 Evidence and proof

The philosophical concept of evidence is one of those present in the stormy philosophical debate about the theory of knowledge for ages. Depending on the philosophical viewpoint chosen, the concept has been differently defined and even confused with other terms. The huge variability of meanings in English of this term, that is, its polysemic nature, unlike its clearly philosophical meaning in Spanish, leads to an outright rejection of using it with such a wide meaning. However, the frequent use of that term in written forensic science in Spanish should be acknowledged and then it can be considered as a neologism (see Lucena, 2016, pp. 147–148 for further details).

In the philosophy of being, it is explained that beings become apparent in the presence of intelligence and such a manifestation causes an act in the intelligence called 'evidence'. It is an intelligible property of beings. That property is the own manifestation of being and it can be described by saying that what is known becomes apparent to our intelligence so clearly that it asserts a judgement in a natural way. As an example to facilitate understanding of its meaning, we can focus on its first meaning: the one linked to sight. What is known by sight has the property of evidence, that is, it is evident. It is not possible to deny what is seen because intelligence is lit by the evidence of a sensible experience. It is a common saying that 'it is not possible to deny the obvious', but it is also certain that we can be reluctant towards it and refrain from making judgements because we are free.

What applies to sight can be applied to the other senses, although it is not only those knowledge faculties that allow us to perceive evidence. Evidence perceived by senses is called 'immediate', but when we use deductive reasoning, the logical conclusion from the premises is perceived by our intelligence as evidence, in this case intellectual, and in a mediate way. The clarity in the intelligence confirming the correspondence between what is thought and what is real (or, where appropriate, what is just merely thought) in a proof has the same strength as that in a sensible experience.

It does not make any sense to speak of degrees of evidence because it refers to the highest possible degree of clarity by which intelligence gets to confirm the correspondence between what is thought and reality (or what is merely thought). However, in specialised scientific literature about evaluation of evidence, the term 'strength of evidence' is often used. This term refers to how probable findings observed by an expert are, given that each of the two exclusive hypotheses supported by parties in a criminal proceeding are true. This is the definition of a likelihood ratio, considered nowadays as the scientifically correct way of evaluating evidence by forensic experts. Therefore, what was earlier said has to do with this similar question: what does data tell us when it is evaluated as evidence? (Royall, 1997, p. 4).

For example, it is common to say in Spanish that the finding of combustion accelerators at the seat of a fire could be evidence of an arson attack. The term 'evidence' is used, in this case, as

a property of an action (as proof), not to designate the action itself or the specimen collected. Evidence is said of a possible ill-intentioned behaviour, which could explain the finding of the combustion accelerator at the crime scene.

We also say that an object seen by sight is evident for our intelligence, but we do not call the object seen 'evidence', but a property of that object for our knowledge faculties.

Finally, when we say that "2 + 2 = 4", it is evident that we are denoting a property of the result reached by the deductive reasoning used.

Evidence is, in the final analysis, what allows intelligence to be certain. Intelligence rests when the strength of the correspondence between what is thought and reality (or what is merely thought) reaches such a degree of clarity (evidence) and not any other.

The term 'evidence' in a Spanish-English English-Spanish dictionary of legal terms such as that of Ariel (Alcaraz et al., 2001) – probably the best dictionary in Spanish – is simply non-existent in the Spanish-English section (translating literally the English term, just looking for *evidencia*).

These are the meanings of 'evidence' in the *Dictionary of the Royal Academy of Spanish Language*: (1) undoubted clear and plain certainty; (2) determinant proof in a proceeding. The first meaning identifies evidence and certainty (it is not philosophically consistent) and the second has a strict legal sense.

However, the English term is very polysemic and therefore it deserves a specific study.

The Ariel dictionary translates 'evidence' for Spanish readers as 'testimony, proof, documentary evidence, means of proof and signs'. The English-English dictionary of Longman (ed. 1987) says the following: "something such as a fact, sign or object that gives proof or reasons to believe or agree with something" and also "the answers given in a court of law".

In the Internet Oxford Dictionary evidence is defined as the available body of facts or information indicating whether a belief or proposition is true or valid.

In the Free Online Dictionary on the Internet it is said that evidence is "a thing or things helpful in forming a conclusion or judgement, something indicative, an outward sign, the documentary or oral statements and the material objects admissible as testimony in a court of law". Also, it is defined as "ground for belief or disbelief, data on which to base proof or to establish truth or falsehood, or a law matter produced before a court of law in an attempt to prove or disprove a point in issue, such as the statements of witnesses, documents, material objects, etc."

From a legal perspective there are some adjectives linked to the English term 'evidence' that are especially relevant for our purpose:

"Direct evidence: evidence (usually the testimony of a witness) directly related to the fact in dispute."

"Circumstantial evidence, indirect evidence: evidence providing only a basis for inference about the fact in dispute."

“Corroborating evidence: additional evidence or evidence of a different kind that supports a proof already offered in a proceeding.”

“Hearsay evidence: evidence based on what someone has told the witness and not on direct knowledge.”

In short, we can distinguish the following English meanings by collecting synonyms in that language:

a) proof, grounds, data, demonstration, confirmation, verification, corroboration, authentication, substantiation. For example: There is no evidence to support this theory.

Here evidence means proof. Though it is possible to distinguish in English between evidence and proof by defining the first as a piece of information that allows us to make an inference, and the second as one that allows us to verify such an inference, a synecdoche is created in practice. The part (evidence) comes to mean the whole (evidence and proof).

b) sign(s), mark, suggestion, trace, indication, token, manifestation. For example: The police said there was no evidence of a struggle.

Here the meaning of clue, sign, trace, mark, vestige, indication, etc. prevails.

c) testimony, statement, witness, declaration, submission, affirmation, deposition, avowal, attestation, averment. For example: Forensic scientists will be called to give evidence.

And in this last sense, it means testimony.

It has already been mentioned that the terminological variety of proof is described by specialists as chaotic and inconsistent, both in the legal doctrine and case law. In a criminal investigation the court is interested in determining the relevant facts to make a decision, but the court cannot prove the facts that really happened but statements on those facts. So Gascon says that “to assert that an explanatory fact is proved or it is a proof means it has been verified, its truth has been checked” (Gascon, 2010, p. 76).

Legal literature has made a distinction between two different contexts in which evidence has to be used: discovery and justification. In the former, evidence consists in discovering or knowing facts that make statements of those facts true, while in the second, evidence consists in justifying the truth of statements related to those facts. While they could be interchanged regardless of context in English, it would only be true in the context of justification in Spanish.

Additionally, according to Gascon again (Gascon, 2010, pp. 76–88), the term ‘evidence’ has different meanings within the legal-procedural vocabulary: means of proof, evidential proceedings or results of evidential proceedings. As means of proof it refers to those things that allow us to know the relevant facts of a case (cognitive function): witness testimony, submission of documents, expert reports, judicial identification, etc. As results of evidential proceedings it refers to the results obtained through means of proof (justification function): the verified explanatory fact describing the controversial fact, for example: ‘there was a weapon at the defendant’s home’, ‘Mr. So-and-so made a phone call announcing the location of a car bomb’, etc. As evidential proceedings, it links these two mentioned meanings (cognitive function): means of proof and a verified assertion on facts. It is a confirmation or an

inference by which, starting from means of proof, finally relevant facts are known and enable a proper decision-making process.

Evidence as a result of evidential proceedings is considered to be the more accurate meaning to the function pursued by evidence in court: to prove the truth of a statement in relation to a relevant fact to make a decision (justification function). The above-mentioned three different meanings are expressed in Spanish using the same word: *prueba*. It is the same in English, although there are two words to designate the same thing: 'evidence' and 'proof'.

As evidential proceedings we can find in literature the following adjectives to distinguish different types of evidence among others: direct, indirect, circumstantial, presumptive, critical, historical, demonstrative, inductive and deductive. Scholars observe differences in meaning when those terms are used by legal doctrine and case law. Perhaps the most common distinction is made between direct and indirect evidence, but we should note that, in this case, Gascon has found three criteria on which to base such a distinction:

a) The immediacy criterion to put facts to the proof or to verify statements of those facts. Direct evidence consists of the immediate empirical confirmation of the statement being proved, and indirect evidence allows us to reach facts put to the proof through another or other facts by means of inference. So, evidence related to past events is always indirect. Only immediate observations of facts put to the proof are direct evidence.

b) The inferential type criterion: if deductive, it will be direct, if inductive, it will be indirect. With the former, deductive logic rules are used, with the latter, probabilistic laws and generalisations are used instead. We can achieve necessary results with the former, but only probable results with the latter.

c) The predominant criteria in both legal doctrine and case law, according to Gascon, are the following:

> c.1) Direct evidence: the fact to be proved appears directly and spontaneously, without mediation and reasoning, from a means or source of proof; it is said that by itself it is able to base the judicial conviction on that fact. It is the same as historical evidence, therefore, as testimony or documentary evidence.
>
> c.2) Indirect evidence: the fact to be proved does not appear directly and spontaneously from a means or source of proof, but it needs reasoning; it is said that by itself it is not able to base the judicial conviction on that fact.

Gascon argues that these criteria are an example of a certain confusion among the concepts of evidential proceedings, means of proof and evidence *in strictu sensu*. The assimilation between the historical and direct evidence is, certainly, false, because historical evidence does not always refer directly to the fact to be proved. The distinction between direct and indirect evidence referring to the evidential proceedings is inconsistent because in both cases the same kind of inference is made. A testimony can only prove that a witness has said something, but not the fact referred to. Finally, the distinction makes sense when considering evidence as a result of evidential proceedings (as a verified explanatory fact) because the adjective 'direct' means that evidence is directly related to the fact in dispute, which we want to prove and on which a judicial decision is dependent, and the adjective 'indirect' means just the opposite.

Gascon argues that the most convincing criterion for distinguishing evidential proceedings is the possibility of observational empirical verification, and if that is not possible, the kind of inference used: deductive or inductive. That's why the former proceeding is called 'direct evidence', and the latter 'inferential evidence'. Within inferential evidence, deductive and inductive or indirect evidence may be distinguished.

So, here is the distinction proposed by Gascon based on a verification logical framework:

Direct evidence: factual knowledge proceeding (or explanatory fact verification) based on observations made directly by judges.

Deductive evidence: factual knowledge proceeding (or explanatory fact verification) based on a deductive inference from other verified assertions.

Indirect or inductive evidence: factual knowledge proceeding (or explanatory fact verification) based on an inductive inference from other verified assertions.

Finally, Gascon distinguishes the following evidence types *in strictu sensu*: (i) considering the evidential proceeding type by which evidence has been obtained; (ii) considering that either it constitutes the aim of the evidential proceedings or it is used as an argument to prove other explanatory facts; (iii) considering that either it is by itself enough to base judicial decisions on or not.

Distinction (i) leads to direct verifications or confirmations, conclusions and hypotheses. Verifications or confirmations are observational statements and the results of direct evidence. Conclusions are the results of deductive evidence. And finally, hypotheses are the results of indirect or inductive evidence.

Distinction (ii) leads to distinguishing pure and simple evidence from evidential statements. The former are explanatory facts as results of evidential proceedings, and the latter verified explanatory facts used as premises to prove other explanatory facts. The evidential statements used to prove others are called *elementos de prueba* (items of proof). The evidential statements of direct or circumstantial evidence proceedings are called *indicios*.

In distinction (iii), incomplete evidence usually contributes to the judicial decision as signs.

In legal doctrine and case law it is usual to identify full and direct evidence as well as incomplete and indirect evidence or signs.

4.1.3 Science, belief and faith

A brief explanation of the topic can be found in (Lucena, 2016, p. 148).

At this point one wonders if faith is always free, as St Thomas Aquinas did. Aquinas distinguishes between free faith and needed faith. For an act of faith to be carried out, the intervention of direct will to determine consent is needed and sufficient. Will, however, is not always free. Reasons may be so convincing that our will cannot resist them. In this case, faith is forced by the evidence of signs. It can even be possible that the difference between both faith types mentioned can be reduced to the point of cancellation: an act of faith, free in principle, may not be such in an individual subject due to habits or social pressure. In contrast, a forced

act of faith will not be such when a margin of freedom persists (Vernaux, 1971, pp. 140–141). And this is also applicable to a merely human faith.

4.1.4 Elucidation of terms: certainty, opinion, probability, evidence and belief

Most of the content of this section has been published in (Lucena, 2016, pp. 149–152) under almost the same heading (with the exception of having added 'subjective' to 'probability').

The Bayesian inference approach enables the combination of probabilities based on data (also called 'objective' or 'frequentist probabilities') with subjective probabilities (assigned by considering experience, knowledge and information) (Aitken et al., 2004, pp. 21–22). On the other hand, Shafer developed a theory of belief based on some ideas provided by Dempster regarded as an extension of the theory of subjective probability.

Shafer criticises the subjective Bayesian probability, emphasising that it is incapable of distinguishing a lack of belief from disbelief, but what is relevant here is not the development of the Shafer theory but the gradual consolidation in the scientific literature about evidence evaluation by forensic experts of the similarity between the concept of belief and subjective judgements, on the one hand, and the measurement of uncertainty of such subjective judgements (statements and propositions) and the measurement of belief of such beliefs (the same statements and propositions), on the other. It is an ambiguous way of understanding the term 'belief' that may cause, in our view, misunderstandings in distinguishing well between the verbs 'to know' and 'to believe'.

When probabilities are thought of as personal degrees of belief – as happens in the Bayesian inference school – we talk about 'strong' and 'weak' beliefs or 'certain' or 'hard' and 'uncertain' or 'soft' evidence (Taroni et al., 2006, pp. 1, 19 and 41). The words used in the Spanish translation have been literally translated from English (*creencia* for belief, and *evidencia* for evidence). In this case, to make a good English-Spanish translation it is important to strictly keep the sense of words in the original language (English). The qualification as 'strong' or 'weak' means that belief, as confidence, is related to more or less uncertainty. Therefore, what is strong or weak is the degree of uncertainty in the truthfulness of the judgement or, in other words, the credibility that a person has about it. Belief as judgement is stating without further ado, and it is not adjustable. Then, what is right – from the epistemological perspective – is to speak about strong and weak credibility regardless of which language is being used.

As for the possibility of an evidence being certain (or hard) or uncertain (or soft), the term is being used meaning proof. In this sense, the very statement that is verified in evidential proceedings, as such a statement, that is, as judgement, is capable of being certain or uncertain. While in the specialised English literature on Bayesian inference in forensic science the term 'uncertain or soft evidence' or 'certain or hard evidence' is used, that's how it should be understood in Spanish. Summarising, 'evidence' is proof but, in turn, the term 'proof' is stated here in the sense of 'evidential statement'.

On the other hand, the term 'certain' or 'hard' evidence is related to understanding that the proposition conditioning evidence whose probability we are interested in is considered to be true, and the term 'uncertain' or 'soft' to the understanding that the above-mentioned conditioning proposition has a probability of less than one of being certain. In the first case, if

we want to know the posterior probability of propositions we will apply Bayes' Theorem. In the second case, Jeffrey's rule.

4.1.5 More on elucidation of terms: uncertainty, determinism, indeterminacy, cause, principle, condition and occasion

The concept of uncertainty can get to subsume that of indeterminacy making up a metonymy. The specialised scientific literature distinguishes between epistemic and random uncertainty. The former is linked to a lack of knowledge or information, and the latter to an observable property in nature. The former is understood as the degree of conviction reached by a person on the truthfulness of a statement, the latter as the effect of the deterministic principle of causality considered evident in observing nature.

Those who have supported a scientific determinism, the paradigm of classic mechanic, have confused the terms 'uncertainty' and 'indeterminacy', giving a meaning to the former outside its epistemological nature, to be a state of mind after making an act of understanding. Uncertainty exists only in the mind. Scientific determinism is said of the real world understood as something outside the mind, the material universe.

When Heisenberg formulated his famous uncertainty principle, he brought into question inside the microcosm the classic mechanistic thesis of the principle of causality.

When an individual knows something real, he/she can be in a mental state of uncertainty as to whether the known phenomenon belongs to the macrocosm or microcosm. In both cases the uncertainty may be epistemic, random or both.

In experimental sciences, the concept of cause used is linked to a temporal priority. If we consider types of causes different from the efficient cause, like material or formal causes, it is possible to understand, for example, that a human being is formed when the soul and body are joined, and in this kind of union the temporal difference makes no sense. Another example in which the temporal sequence has a different order from that between observable phenomena in nature is the final cause. Conscious human activity always has an aim called the 'final cause'. Without that aim there would not be an explanation for the performance of a determined action, action that would be linked to an efficient cause submitted to a chronological temporal order. Speaking about chronology, in this example the final cause is temporally prioritised, which explains the existence of an efficient cause acting and producing its effect afterwards.

The existence of causality in reality, such as it has been defined, is evident for any intelligence. If this is asserted it is because its foundation is being, which is, and, consequently, it can cause. There is a known Aristotle-Thomas aphorism that states: *agere sequitur esse*, that is, action follows being.

It is only possible to understand well the principle of causality of the philosophy of being starting from its concept of being: the notion of being is not a genus; it is not possible to add any difference not previously included (see Lucena, 2016, pp. 154–155, for further details).

To distinguish the concept of cause from the remaining mentioned concepts in the title of this section, principle, condition and occasion (see Lucena, 2016, p. 155). Likewise, some clever

comments coming from (Arana, 2012) regarding determinism, indeterminacy and chance have been included in (Lucena, 2016, p. 156).

4.1.6 Statistical evidence and confidence (degree of belief)

Royall explains with simple and clear examples the difference between evidence (statistical evidence) and confidence (degree of belief) (Royall, 1997, pp. 13–16).

In essence, the key to distinguishing them is to wonder about those questions after observation:

- What do I believe, now that I have this observation?
- What does this observation tell me about A versus B? (How should I interpret this observation as evidence regarding A versus B?)

The first question is related to degrees of belief, the second to statistical evidence. But a new problem rises if we understand 'degree of belief' as 'confidence'.

It seems unfortunate to explain the same concept (subjective probability) by means of different words, even more so if they are polysemic. It is not good for the sake of clarity to confuse the meanings of the verbs 'to know' and 'to believe', interchanging them again and again. If the verb 'to know' is easier for us to interpret according to the same sense in the ordinary linguistic practice, it is not the same thing regarding the verb 'to believe', which fluctuates, that is, sometimes it means 'to know' and sometimes 'to believe'.

The first question has to do with an act of knowledge (in spite of the fact that the verb 'to believe' has been used). When it is believed that an act of knowledge is uncertain, that is, when certainty has not been achieved, it is possible to speak about 'degrees of uncertainty', and this is exactly what is meant when we speak about 'degrees of belief' because we wonder which knowledge could be achieved of a problem when the observation has been made. That's why we are at the probability scope, either expressed in terms of degrees of credibility (not properly degrees of belief as explained in (Lucena, 2016, pp. 151–152) and where it is said that one way of dealing with uncertainty is to understand it as a belief function).

Intelligence may consent to a judgement by being determined by the known object (it has been designated as intelligence or science depending on the immediacy of the known object to whoever knows it). However, intelligence may consent to a judgement through one's own free will. The assent may be made with more or less firmness and, therefore, one could speak about certainty, opinion and doubt in both contexts. In the science scope we may achieve physical certainty, though in experimental sciences it coexists with opinions, conjectures and doubts. In the belief scope we may achieve certainty-belief (act of faith), but opinions, conjectures and doubts understanding belief as credibility.

4.1.7 Certainty, probability and entropy

The probability and information theories have developed a mathematical representation of the state of mind of certainty regarding the occurrence of an event or the truthfulness of a statement. The former measures it with two opposing values: 0 and 1, where 0 means negative certainty and 1 positive certainty of the occurrence of such an event. The latter

measures it by means of an entropy value equal to 0, with that number representing both positive and negative certainty.

With direct measuring the monotonic principle should be followed, that is, the measurement value must be a direct mathematical relationship with higher or lower uncertainty. Therefore, the bigger the numeric value the more uncertainty we will have, and slight changes in uncertainty should result in slight changes in measurement values.

Probability measures the credibility degree that anybody has in the occurrence of an event or the truthfulness of a statement. Such a degree of credibility can be based on a frequentist concept of probability when the nature of the fact allows us to assign that probability, or on a subjective concept of probability which besides including the classical concepts of probablity also allow us to calculate probabilities of unrepeatable facts. The degree of credibility is related to the uncertainty measurement in an indirect way because it needs a previous mathematical transformation. The maximum uncertainty is represented by means of the equal probability among all the possible states – let us say n, of a random discrete variable. Its numeric value is 1/n. As n represents a number of possible states in the real line, mathematically there can be up to infinite equally probable states, that is, the ways by which probability (or the degree of credibility) represents the maximum uncertainty are infinite.

If we try ideally to show the maximum uncertainty irrespective of the number of possible states related to the occurrence of an event or the truthfulness of a statement, a good solution can be found in the entropy theory. It is intuitively easier to relate any degree of uncertainty with respect to a possible maximum in a linear way than to get it by means of any other procedure involving a further mathematical transformation, as simple as it could be. This idea underlies the selection of the concept of normalised entropy, coming from Shannon's information theory, as being particularly attractive to our aim.

Let us go more deeply into the mathematical concept of entropy according to Shannon's information theory (1948). We have a probability distribution $P = \{p_1, p_2, ..., p_n\}$. We define the entropy of the distribution of probabilities P by: $H(P) = \sum_{i=1}^{n} p_i \times \log(1/p_i)$ for discrete probabilities, and $H(P) = \int p(x) \times \log(1/p(x))\,dx$ for probability density functions. It can be shown using the Gibbs' inequality for discrete probabilities that the entropy function has a minimum of 0 when one of the probabilities is equal to 1, and the remainder equal to 0, and a maximum value of $H(P) = \log(n)$ when all the events are equally probable, that is, 1 / n. For this very reason, the maximum value log(n) is used as a normalisation factor.

Another way to define entropy is in terms of expected value. Given a discrete probability distribution p_i, with $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$, or a continuous distribution P(x), with $p(x) \geq 0$ and $\int p(x)dx = 1$, we can define the expected value of an associated discrete set $F = \{f_1, f_2, ..., f_n\}$ or a function F(x) by: $< F >= \sum_{i=1}^{n} f_i p_i$ or $< F(x) >= \int f(x)p(x)dx$. From this definition entropy can be defined as the expected value of the information of the distribution I(p), that is: $H(P) =< I(p) >$.

Entropy and probability are formally related to each other by means of the shown mathematical equations. They are interdependent concepts. This formal relationship between them allows those concepts to be used to measure uncertainty.

Courts have to evaluate statements about facts given the background information of the case and evidence available. Their a posteriori odds are based on two mutually exclusive propositions, according to the alleged statements from the parties. In this framework, a posteriori probabilities are linked to binary discrete variables (the statements can only be true or false when compared with the reality of the facts). Though the odds version of Bayes' Theorem does not imply that propositions have to be exhaustive, the fact is that the context of penal proceedings determines generally the exhaustiveness between propositions. In fact, in most cases where evaluative conclusions may be given (those in which likelihood ratios are assigned as the strength of the evidence), the alternative proposition is probabilistically complementary to the earlier proposition (Robertson et al., 1993, pp. 471–472; Taroni et al., 2010, pp. 62–63).

Inside this framework an ideal method of numeric representation of uncertainty by the court about the propositions from the parties could be considered, due to particular properties making it easily intuitively understandable, to represent uncertainty numerically with respect to a possible maximum in a monotonic way. In this sense it is ideal to use the interval [0, 1] to associate the value 0 with the absence of uncertainty (state of certainty), and the value 1 with the maximum uncertainty. So no additional mathematical transformation to represent uncertainty would be necessary, because the used number would already do it directly. While the maximum uncertainty is numerically represented by the value 0.5 for a binary discrete variable, the normalised entropy represents it with the value 1. Likewise, while certainty is represented by probability by means of two values, 0 and 1, the normalised entropy only needs one value, the figure 0.

One of the main properties of the entropy function H(p) is that it is maximised when the probabilities of all possible states of the variable are equal. The notation log(n) represents such a maximum value, with n being the number of states. The higher the number of states of a variable, and when all the possible states are equally probable, the higher the entropy value will be.

When the entropy of a discrete variable of n possible states is calculated, a normalisation of its value with respect to the maximum value of the entropy function is possible. So, the value of the function is expressed in a monotonic way and in the interval [0. 1]. The resulting value can be interpreted as a direct measure of uncertainty with respect to its maximum value, which is just what we were looking for.

From a merely intuitive viewpoint, what advantages or disadvantages does thinking in terms of probability or normalised entropy to measure uncertainty have? It seems clear that the probability concept has several centuries of history, has been successfully used for a long time in all kinds of sciences, has been enriched by different interpretations and theories, and it is only possible to establish its scientific relevance. However, it is acknowledged that probability does not measure uncertainty in a monotonic way. This limitation brings an interpretative complexity to its numeric values, increased by the apparent monotonous increase or decrease of the measurement of uncertainty derived from the possible probability value range (1st axiom of the theory of probability). Indeed, monotony only has to do with the adjustment of

credibility (interpreting probability as degree of credibility) or the magnitude of relative frequency (interpreting classically probability).

The concept of probability as a measurement of uncertainty of the occurrence of an event has been mathematically defined within the classic axiom in such a way that it represents certainty by means of two different numbers: negative certainty (false) and positive certainty (true). This is not the case with the normalised entropy. Is there any advantage in distinguishing between positive and negative certainties? We think so. Both are certainties and from this exclusive aspect they are not different from each other, however the associated numeric probabilistic value provides added information (it qualifies the kind of certainty). Normalised entropy only measures uncertainty and therefore it only needs a zero value to express its absence.

The simple numeric value of the normalised entropy to express both certainty and maximum uncertainty, and its numeric direct relationship to a monotonic measure of uncertainty, allows it to have an advantageous position against how probability measures it.

The axioms of the probability theory have made the development of an unchallengeable scientific value of mathematical logic possible. The distinction between the negative and positive certainty states, and the representation of maximum uncertainty by infinite values between [0.5 and 0] have been useful for graphic interpretation of probability in a very intuitive way in different dimensions.

Probability only measures directly uncertainty when it is 0. When the probability of an event is 0, that event is called an 'impossible' event and there is no uncertainty at all, therefore if the uncertainty value is measured in the interval [0 1], the value 0 means exactly what we are looking for with such a figure: there is no uncertainty at all.

4.1.8 Probabilistic explanations and conjectures

Gascon (Gascon, 2010, p. 104) says that from some evidential statements and certain causal regular rules (usually simple lessons of experience) a hypothesis about facts is reconstructed to explain evidence better than any other. As those causal regular rules are probabilistic and, in addition, some evidence has been mediately obtained, the explanatory hypothesis of those facts can only be probable. Therefore, the representation of fact of a sentence is not sure. *In strictu sensu*, it can only be stated that indirect evidence concludes by making an assumption, a statement that we consider to be true, but we do not know if it is true for certain. For further details, see (Lucena, 2016, pp. 142–145) under the heading: "The role of logic of propositions".

4.2 International norms and guides on interpretation of results in forensic science

We find ourselves before a developing task, that is, there is no formal European standard on evaluative conclusions for forensic reports yet, but a process has been initiated within a Committee for European Normalization (CEN 419) called 'Forensic Science Processes'.

There are no known analogue examples in other places in the world, but there could be an exception with the recent Australian standard on forensic science, which focuses on the interpretation of results of forensic analyses (AS 5388.3-2013). Nevertheless, its scientific perspective are very far from that being considered in the European documents.

Otherwise, in other more general standards related to forensic science such as the G-19 from ILAC (International Laboratory Accreditation Cooperation), some lines on conclusions of forensic reports have been written. Though they have done that under the wide perspective of the document wherein they are, they ask to be respectful of principle b) of the forensic evaluation detailed in the next section.

We would have to go back to the year 2008 to find a European proposal of a standard on the matter at hand: that developed by the AFSP (Association of Forensic Science Providers; see AFSP, 2009) for England, Wales and Ireland.

And referring to a widely European scope, the development of the project titled 'Development and implementation of an ENFSI standard for reporting evaluative forensic evidence' within the Monopoly 2010 programme of ENFSI, financed by the ISEC (Prevention of and Fight against Crime) programme of the European Commission could be highlighted. This project finished in December 2014 with the delivery of a leading guide for ENFSI about evaluative conclusions. This document is considered to be the basis for the development of the European standard in this topic, which the European Committee of Normalization 419 intends to write in its part 3, the same as the topic division of the Australian norm AS 5388. Through decision 42-2015, CEN/TC 419 has transferred all its work items to ISO/TC 272 Forensic Sciences/WG3 Processes, under the Vienna Agreement, which lays down the conditions when envisaging the transfer of CEN work to ISO. Specifically, CEN WI 00419003, 'Evaluation and interpretation of the results of forensic science examinations and analysis in the context of the case', has been related to the ISO future new work item 'Forensic Analysis. Part 3: Interpretation'.

4.2.1 Principles of forensic science evaluation

It seems to be necessary to start delimiting the scope of the document that is being written in the framework of the Monopoly 2010 programme because, indeed, it is not trying to propose a future standard for all kinds of answers from a forensic science laboratory at the request of its clients. It focuses its attention on a determined type of conclusions called 'evaluatives'.

It is common to quote the work titled *Interpreting Evidence. Evaluating Forensic Science in the Courtroom*, of B. Robertson and G.A. Vignaux, published in 1995, as the main explicative work of the concept of evaluative conclusion used in this thesis in the forensic science scope.

And in trying to precise such a concept to the maximum within the above-mentioned scope, the document that is being written by the ENFSI research group refers explicitly to the paper written by I.W. Evett, G. Jackson, J.A. Lambert and S. McCrossan in *Science & Justice*, Volume 40(4), in 2000, titled 'The Impact of the Principles of Evidence Interpretation on the Structure and Content of Statements'.

The quoted paper includes the following three rules to be taken into account in order to make a forensic science evaluation:

a) Interpretation of scientific findings is carried out within a framework of circumstances. The interpretation depends on the structure and content of the framework.

b) Interpretation is only meaningful when two or more competing propositions are addressed.

c) The role of the forensic practitioner is to consider the probability of the findings given the propositions that are addressed, and not the probability of the propositions.

These three rules describe the essence of an evaluative conclusion (see also Evett et al., 1988). In practice, they are related to forensic comparisons between questioned and unquestioned samples, the former obtained at the scene of the crime, from the suspect or victim, and the latter as a consequence of police records of suspects or undoubted samples expressly recorded to make the forensic report. The aim of these comparisons is for the court to evaluate the extent to which the questioned samples could be attributed to the corresponding suspects bearing in mind the propositions on this issue supported by the parties in the proceeding.

Therefore, it may be considered that DNA profiles, fingerprints, shoemarks or tyre tracks, handwriting or handwritten signatures, crystals, paint, fibre, ballistic fingerprinting or speaker comparisons, and others similar to them, are all included in the scope covered by the document, that's why it is so relevant.

The existence of approximately 30 years of scientific literature specifically related to forensic comparisons with character identity in the above-mentioned disciplines should be emphasised.

4.2.2 Typology of conclusions in forensic reports

The ENFSI Guideline accepts that a laboratory may respond to an applicant for a forensic report with a diverse typology of conclusions because this is the logic nature of answers according to the raised questions. We can distinguish between technical and factual, investigative and operative, intelligence, and evaluative conclusions. Though only evaluative conclusions are dealt with in the document, it has to define all the possible types as well.

Factual or technical conclusions are those that only require a technical interpretation of the results. For example, when a specialist measures the sound pressure emitted by a noise source from a certain distance with a sonometer, we can only interpret the meaning of the datum obtained in its measure and associated uncertainty. And that is a merely technical interpretation, as previously mentioned above.

Investigative conclusions require a context wherein the crime investigation is in what can be called an explicative phase. There is a need to formulate hypotheses to explain the already known facts, which deserve attention. For example, when an arson attack is investigated, an expert tries to get an idea of its cause, and its possible spreading according to what has been observed on the ground, and, of course, his/her knowledge and experience.

Intelligence conclusions allow police investigators to link cases, events and situations to inform strategical or operational decisions.

Evaluative conclusions evaluate the forensic findings following the principles of forensic science evaluation already quoted in an earlier paragraph.

Therefore, it is possible to find factual, investigative, intelligence or evaluative conclusions according to the questions raised to the laboratory. That distinction has a merely logic nature and, consequently, they are brought up to an abstraction level compatible with any judicial system.

4.2.3 Determination of propositions in evaluative conclusions

The document emphasises the need for a close collaboration (or with the parties in the Anglo-Saxon system) for the expert to determine propositions to bear in mind in his forensic evaluative examination so that he can provide the best possible information for the court to make decisions (Taroni et al., 2013, pp. 467–470).

The evaluative reports are dependent on the context in a specific criminal case – which only the judicial authority is able to know completely in Spanish penal proceedings – whereby the glossary defines a key concept: conditioning information. Guidence note number 2 of the document is focused on a novel concept for many forensic experts, that of proposition hierarchy. The basic idea is to acknowledge that propositions formulated by the expert in an evaluative report may mean that they have to be described in the context of an activity instead of doing it as a mere coming or origin from questioned samples. For example, the absence of fibres on the seat of a vehicle wherein there is a record of the occurrence of a struggle between an aggressor and his/her victim may be significant as a great number of them could be expected in those circumstances (nature of the struggle, time passed and so on). In this sense, the concepts of transfer, persistence and background levels are especially relevant.

4.2.4 Pre-assessments in evaluative reports

The concept of pre-assessment is not something simply related to the Anglo-Saxon practice of considering the submission of forensic reports made by laboratories to be part of the market, something like a criterion of subsistence economy in a judicial system wherein the forensic evidence practice is not ordinarily free of charge.

The concept of pre-assessment tries to avoid forensic experts making evaluative forensic reports in previously unexplored contexts, that is, wherein the strength of evidence expressed by numeric likelihood ratios does not have a predictable reference.

Also, pre-assessment avoids formulating propositions led by data, though it is acknowledged that in the very first moments of an investigation it is not possible to avoid it.

The above-mentioned reasons advise us to include the practice of pre-assessment in the document, because it is a document that has to fulfil the exigencies of all the European judicial systems.

Pre-assessments are established when propositions have to be formulated at the activity level because mechanisms of transfer, persistence and background levels are considered to be relevant in determining the strength of the scientific findings as evidence. That is explicitly written in guide note number 3 of the document.

4.2.5 LRs faced with lack of data or insufficient data

This subject was one of the most commented upon subjects by the ENFSI members once they received version 2.7 of the document. Many of them argued that it was not possible to apply the LR approach in their fields due to a lack of databases or suitable probabilistic models for their casuistry. The research group proposed a change wherein it is made clear that the

document emphasises that LRs based on data are the ordinary way in which the evaluative reports should be concluded, though before absence or not enough data verbal expressions of non-numeric LRs can be used based on personal knowledge. The reason why this is so is that the LR mainly expresses a logical framework for evaluating scientific findings as evidence.

The document also shows that it is possible to use verbal expressions of LRs from numeric LRs by means of using corresponding scales, but underlines that it is not permitted to assign the numeric LRs of the scales from verbal expressions of LRs before the absence or not of enough data.

In any case, the laws and theorems of the probability theory have to be respected when any probability is assigned to a proposition. The work of O'Hagan and others titled *Uncertain Judgements: Eliciting Experts' Probabilities*, published by Wiley in 2006, is recommended on this matter (see also Taroni et al., 2016).

4.2.6 Application problems of the ENFSI Guideline in the judicial scope

It is evident that the interpretative changes mentioned regarding conclusions of forensic reports cause firstly a serious problem within each laboratory, whereby a specific training plan is required to be able to adapt to the requirements of the predictable future standard, and an associated problem – perhaps still more serious than the quoted one – in the judicial scope.

The project group has made a road map allowing each laboratory to apply a training route to reach an effective implementation of the future standard. Without doubt, it is the first target that everybody has to reach.

The difficulties for the judicial fora in understanding the predictable standard will be reduced, likewise, with specific training plans for their bodies. There have been experiences in that sense in Europe: Sweden, the United Kingdom, Belgium, Switzerland, Holland and Spain.

4.2.7 Example of application of the ENFSI Guideline in forensic acoustics (voice comparisons) – see APPENDIX IV (p. 160)

The case example is outlined on the left-hand side of the pages while the relevant section of the ENFSI Guideline and additional information are on the right-hand side.

4.3 Moving from the individualisation to the likelihood paradigms in forensic science in Spain

Chapter 6 has its own story. May we apologise to readers for the biographical character of the following lines. This doctoral thesis would never have been submitted without the personal experience lived by this author after working in the same official laboratory for more than 26 years.

The University Institute for Research in Police Sciences (IUICP) of the University of Alcalá de Henares approved a research project in 2009 titled 'The Evidential Value of the Official Forensic Science Report: towards a legal amendment from a higher jurisprudential acknowledgement', with this reference: IUICP/PI2009/002. The institute was created in 2005 after an agreement was signed between the Secretary of State for Security of the Spanish Ministry of the Interior and the university mentioned above. Three main Spanish official forensic laboratories are members of the institute: the Central Laboratory of the Scientific

Police of the National Police (CNP), the Central Laboratory of the Civil Guard (GC), both belonging to the Ministry of the Interior, and the Toxicology and Forensic Sciences Institute (INTCF), belonging to the Ministry of Justice.

The research project was led by M. Dolz-Lago, Prosecutor of the Spanish High Court (SHC), representing the collaboration of the State General Office of the Public Prosecutor in the project, and the last result was the publication of the book *The Scientific Evidence*, whose authors were M. Dolz-Lago (SHC), N. Exposito-Marquez (GC), C. Figueroa-Navarro (IUICP), A. Gomez-Garcia (INTCF), J. Martinez-Garcia (GC), J. Otero-Soriano (CNP), M Santano-Soria (CNP) and S. Valmaña-Ochaita (IUICP), in 2011.

The aim of the research project was to explore which forensic reports carried out by official forensic laboratories nowadays could be considered reliable as documentary evidence. It was not a novelty in Spain because the Spanish Code of Criminal Procedure was amended in 2002 taking into consideration the official forensic reports related to the determination of the nature, amount and purity of drugs seized as documentary evidence in the so-called 'brief criminal proceedings'.

The leader of the project, Prosecutor Dolz, asked official laboratories to answer the following questions (it was approved as a formal decision in the first meeting of the initial research group held on 16 April 2009):

a) updated list of forensic analysis carried out in each laboratory;
b) provisional scientific evaluation of the reliability of each forensic analysis.

Only one official laboratory responded to this request before going to the following project meeting (22 October 2009): the Central Laboratory of the Civil Guard. It was our turn because previously Lt. Col. Montes-Lopez, then our superior officer, had delegated his participation as project partner to us. Eleven pages were written to respond to question b) and the following files were attached and submitted to the members of the research group: the NAS (National Academy of Sciences) 2009 report, the AFSP (Association of Forensic Science Providers) standard and the final report of an ENFSI proficiency test organised by BKA (Bundeskriminalamt) in 2008.

In addition to specific references to the 2009 NAS report, for instance recommendation number 3 encouraging investigation to achieve accuracy, reliability and validity in forensic disciplines as a problem of a general nature, a proficiency test within the ENFSI Drugs Working Group allowed us to emphasise the differences among ENFSI laboratories, making something qualified in Spain, by means of a legal amendment, documentary evidence because the Spanish legislator assumed the above-mentioned properties of the analysis carried out by official laboratories to determine the nature, weight and purity of drugs seized as something already pacifically agreed by the scientific community.

There was a meeting organised by Lt. Col. Montes-Lopez some months later (29 April 2010). It was held in the Headquarters of the Civil Guard to try and bring the two sides closer. A. Carracedo, a well-known Spanish forensic scientist, was invited to moderate a predictably hard debate along with two full professors from the Polytechnic of the Autonomous University of Madrid, Drs. J. Ortega and J. Gonzalez, in favour of our personal position. Unfortunately there was no debate. There were two opposing positions: the legal core research group position expressed by the spokeswoman, C. Figueroa, and the divergent core research group position

expressed by the spokesman, J. Ortega. In spite of the clear support in favour of the 'likelihood' paradigm in forensic sciences given by Dr. Carracedo, the leader of the IUICP project decided to continue according to the initial schedule regardless of the recommendations provided by the guests.

A few months later, the IUICP Board organised a seminar titled 'The conclusions of forensic reports' (24 September 2010) in which Lt. Col. Montes-Lopez spoke about how to defend forensic reports before the court from his long experience. He had just pointed out that the expert has to show the firmness of his convictions about the questions requested in order to avoid the appearance of a lack of confidence in his opinions when we put our hands up and disagreed with that view. Our comment tried to underline the repeated use of the term 'approach' in English scientific literature. In other words, we tried to say that experimental science is not an exact science. The predominant mentality among forensic practitioners who are policemen at the same time is clearly in favour of the prosecution thesis, and they are convinced that their works are really useful if the police investigation ends up in the arrest of those people who committed crimes. Moreover, they think that their works are really successful if their conclusions are categorical or provide probabilities close to one in favour of the prosecution hypothesis.

From then on we decided to leave the research project group and our refusal was quickly accepted. We were replaced by Lt. Col. J. Martinez-Garcia who attended other later meetings and participated until the project's completion.

The following direct quote comes from that book: "The solution consisting of saying that the assessment of the expert and scientific evidence is discretional for the judge sounds like a very insufficient solution. (...) it is not enough to rely only on the 'free assessment' of the tribunal to guarantee that solid science is properly used and correctly interpreted as a basis on which to decide on facts in dispute. What is required for valid scientific evidence to offer a rationale to decide on the facts is a deep and clear judicial analysis of scientific evidence according to reliable standards of evaluation" (Taruffo, 2008).

In spite of this wise recommendation, the authors of the quoted book concluded as follows:

"What scientific evidence could be considered documentary evidence?

- DNA
- Ballistic and tool marks
- Document examinations (not handwriting)
- Forensic entomology
- Fibres
- Fire investigation (only analysis of combustion accelerators)
- Fingerprinting (comparison of fingerprints and palm prints. Developing laboratory procedures)
- Paints
- Gunshot residue
- Soil, plants and fauna
- Toxicology (drugs, already done)

What scientific evidence could not be considered documentary evidence?

- Voice comparison
- Forensic anthropology
- Digital evidence
- Crime scene investigation
- Image technology"

They presented this proposal as something agreed between the legal part of the research group and members of the three main Spanish official forensic laboratories. Among the authors are the heads of two (CNP and INTCF) of the three main laboratories at the time. In order to reach an agreement, several partial meetings were held according to the forensic field just to discuss the questions mentioned above. Forensic experts from each discipline coming from the three laboratories participated in those meetings.

This was the background that explains why we got in touch with two Spanish scholars specialised in philosophy of law, on the one hand, and penal procedural law, on the other, in 2010: M. Gascon and V. Pardo, respectively. The following papers were published as a consequence of a narrow collaboration: Gascon, 2010a; 2010b, pp. 95–106; and Lucena et al., 2011; 2012.

Before the amendment of the Spanish Code of Criminal Procedure by Law 9/2002 on 10th December of Child Abduction, specifically article 788.2, taking into consideration the official forensic reports related to the determination of the nature, amount and purity of drugs seized as documentary evidence, consolidated jurisprudence of the Second Law Court of the Spanish Supreme Court in the eighties and nineties distinguished forensic reports made from official laboratories following official protocols from those made by others. Then a report of the Spanish Judicial Power General Council in June 2002 prepared the subsequent works carried out by the members of parliament both in Congress and the Senate (November 2002). Eventually, the amendment was made as has been mentioned above.

The paper included in Chapter 6, published in the *Journal of Forensic Sciences* in 2012, has the following structure: (1) worldwide overview on criminalistic identification; (2) scientific criticism of criminalistic practice; (3) the Spanish legal reform converting scientific evidence into documentary evidence and a brief comparison between some consequences of this reform and the US Supreme Court case of Melendez v. Diaz; (4) some opinions and possible solutions bearing in mind the good experiences in some countries.

Fortunately, the possible generalised conversion of scientific evidence into documentary evidence has not been carried out in Spain so far. However, this has been mainly due to the strong disagreement among political parties in Spain over approving a new Code of Criminal Procedure (bill of 2013 promoted by Spanish Ministry of Justice).

4.4 Helping judges and juries to apply the Bayesian approach using evaluative reporting

The origin of this other paper also has a biographical character, this time related to how to defend evaluative conclusions before courts and juries absolutely ignorant of the underlying principles of the 'likelihood' paradigm.

The Acoustics Area of the Criminalistics Service of the Civil Guard started to submit evaluative conclusions in voice comparisons from 2004. From 2006 on, the forensic practitioners of this area began to feel the difficulties in explaining evaluative conclusions before courts and juries.

At that time we were the Head of the Acoustics and Image Department and the Technical Director of the Acoustics Area, therefore we personally lived some experiences related to the topic of the article, one of which was especially significant. A lawyer who carefully read our forensic report tried to understand the content, asking a university situated close to her home for help. The result was negative and she asked the investigating judge for the appearance of the forensic practitioner before him and the parties to understand the conclusions before the trial. The judge authorised the appearance and it was held as the lawyer requested. Two hours were needed to carefully explain the methodology used, and to answer all the questions asked mainly by the lawyer and on fewer occasions by the prosecutor. The greater interest shown by the lawyer in the issue was one of the keys to understanding why the prosecution did not obtain a proper benefit from the results of the comparison (all the assigned likelihood ratios in the case at hand were higher than one).

Another outstanding experience was that mentioned in the paper, the forensic report on voice comparison related to the warning of a van-bomb parked at Terminal 4 of Madrid-Barajas Airport in 2006, which eventually exploded. In this case the forensic report was defended before the court and explanations were given by forensic practitioners to make it easier to understand the evaluative reporting. Specifically, the experts tried to give an example of how the members of the court could assign prior odds to multiply by the likelihood ratio submitted in their conclusions in order to obtain posterior odds.

Bayesian networks had been successfully applied in some forensic disciplines up till then. Though the assignments of prior and posterior odds are both, clearly, roles of courts and juries, there can be no doubt that it is possible that experts can help them to do it. It requires using the information that only courts and juries have.

The ENFSI Guideline on evaluative reporting recommends that experts should consider all available, relevant information and, where necessary, request additional information, to establish the key issues (point 3.1). This recommendation could also be applied for the experts to help judges and juries to assign odds in some cases wherein likelihood ratios were given in the conclusions of the expert reports *mutatis mutandis*.

On the other hand, a workshop of the Marie Curie ITN 'Bayesian Biometrics for Forensics', or BBfor2, held in Martigny (France) in December 2011, specifically a presentation submitted by N. Brümmer, an outstanding scientist in speech processing, inspired us to use Bayesian networks to assign prior odds as a tool for helping judges and juries to apply the Bayes' Theorem when evidence is evaluated by likelihood ratios. Of course, the assistance should be understood bearing in mind that such an assignment must be made by judges and juries, but forensic experts may help them to do it according to a solid logical framework such as the Bayesian approach.

In addition, it is interesting to be aware that the Spanish Code of Criminal Procedure distinguishes between the investigative and the trail stages. Depending on the moment at which the assistance is made, Bayesian networks will be able to adapt to the information available and the assumptions made by judges and juries.

The paper included in Chapter 7 contains a particular legal study from the Spanish viewpoint. It is believed that there can be no doubt that the principle of equality of arms is strengthened by applying this new epistemological guarantee in criminal proceedings.

4.5 A new tool for checking the performance of likelihood ratios

When the company AGNITIO S.L. launched the first version of BATVOX (an automatic speaker recognition system for forensic evaluative reporting) onto the market in 2004, nobody had the experience gained by the Acoustics Area of the Criminalistics Service with such a technology and such a procedure to evaluate evidence. That is true because the technology and procedure were transferred by the university research group ATVS, belonging then to the Polytechnic University of Madrid, and that technology was the result of more than seven years of close scientific collaboration between that research group and the Acoustics Area already mentioned. AGNITIO S.L. was a spin-off company of that university at the very beginning, at which point some problems were quickly detected.

The first three real cases investigated by the Acoustics Area using BATVOX 1.0 were unsuccessful. They were related to wiretappings, and the amount and quality of speech of those recordings were good enough for the application of the software and procedure, but the obtained results were far from matching expectations. Previously, the ATVS research group had successfully finished three years of intense activity in wide and specific experimental proofs, even successfully participating in the NIST Evaluations organised by the Department of Commerce of the United States in Washington DC year after year since 2000 and in the first forensic speaker recognition evaluation organised by the NFI (the Netherland Forensic Institute) in 2003.

Nonetheless, AGNITIO S.L. decided unilaterally to check specifically the performance of the likelihood ratios calculated by its system. The company financed a degree project developed by a young telecommunication engineer at the Polytechnic University of Madrid, R. Tapias (Tapias, 2005). The experimental work finished before the summer of 2005 and the project was submitted by R. Tapias to that polytechnic in July 2005.

That job completely changed the situation of crisis caused by the first three forensic reports done in the Acoustics Area, which implied that the experts in that area stopped doing more reports of the same kind and the production line of forensic reports restarted again because the expectations were fulfilled.

Many important things were learnt through such an experimental job with respect to a reliable calculation of likelihood ratios using BATVOX 1.0. Let us call the resulting file once the speech uttered by a known speaker is trained ‘training model’. A training model can be used as a model of a reference population, then it will be called ‘reference model’, and also as a model of a specific person, and then it will be called ‘undoubted model’. Experiments are planned simulating multiple voice comparisons, that is, all test files are compared with all undoubted models generating target likelihood ratios and non-target likelihood ratios. To calculate each likelihood ratio (each test file is compared to each undoubted model) a reference population is needed. The undoubted model selected cannot be a model of the reference population at the same time. All possible combinations are used.

First, an important influence of mismatch in the time between the model trained using the undoubted speech (undoubted model) and the models trained using the speech of the reference population (reference models) was detected. Experiments were performed using the database Baeza_Andaluces, recorded by the Acoustics Area of the Criminalistics Service on the GSM channel. A total of 102 male speakers were available in three well-separated sessions, but only 42 speakers were used. The undoubted models were 60 seconds long and the reference

models 30, 60, 90, 120, 180, and 240 seconds long. Ten 20-second test files were used for each of the 42 speakers. The results obtained were as follows:

a) using a population with reference models longer than the undoubted ones provoked a decrease of the likelihood ratios in general. There was a clear shift of target and non-target curves in the Tippett plots towards the left hand;

b) using a population with reference models shorter than the undoubted ones provoked an increase of the likelihood ratios in general. There was a clear shift of target and non-target curves in the Tippett plots towards the right hand;

c) using a population with models adjusted in time to the undoubted model produced the best performance.

There is an ideal normalisation technique to avoid this mismatch called 'T-norm'.

Second, another important influence of mismatch in the channel between the speech used to train the undoubted model and the speech used to train the reference models was detected. Experiments were performed using the database Ahumada, recorded by the Acoustics Area of the Criminalistics Service in both microphone and landline channels. A hundred male speakers were available in three well-separated sessions, but only 52 speakers were used. All models were 60 seconds long and also all trained from speech of the same session. Eight 15-second test files from two sessions different from that mentioned earlier were used for each of the 52 speakers. The results obtained were as follows:

a) when the undoubted models and test files had the same channel and the reference models had a channel different from the former ones, a shift towards the right of both target and non-target likelihood ratio curves was produced;

b) when the reference models and test files had the same channel and the undoubted models had a channel different from the former ones, a shift towards the left of both target and non-target likelihood ratio curves was produced;

c) when the undoubted and reference models had the same channel and the test files had a channel different from the former ones, the shift was negligible.

d) if all the test files and models had the same channel, the performance was the best one possible. The performance in the microphone channel was better than in the GSM channel.

In the case of a) and b) a more or less narrowing effect between curves was produced as well. This was interpreted as a certain loss of discrimination power of the system.
Other variables such as gender and language produced similar effects to those already depicted.

The clear shifting of target and non-target curves due to some critical factors suggested the idea of checking the performance of likelihood ratios taking into consideration the boundaries of misleading evidence as R. Royall and other authors explained in some papers and books. The paper included in Chapter 8 shows the benefits of drawing limit Tippett plots to detect possible non-calibrated likelihood ratios. Basically, the simplicity of this checking process makes it advisable for forensic practitioners to use.

Chapter 5. CONCLUSIONS

5.1 Some epistemological conclusions

Metaphysics, logic and experimental science are harmonised using a moderate realism based on the gnoseological realism that can be found in the Aristotle-Thomas Aquinas school of knowledge.

Modern philosophy, which can be dated from Descartes on, and which can be characterised as being based on his Principle of Immanency, does not base certainty on evidence but vice versa, and thus truth is based on certainty in the final analysis. This philosophical subjective attitude set accuracy as a prior condition for the possibility of any proper knowledge, but it is not real because it is not possible to anticipate knowledge, to reach an earlier knowledge to all knowledge. To know something is a perfect act and its aim is reached immediately. To reflect is always something posterior, because what is immediate is to know something. Human beings are open to reality immediately by means of intelligence. They do not need the will to do so. Those who defend just the opposing opinion are voluntarists or subjectivists.

The intentionality of knowledge, that is, the fact that the mind and what is known are 'intentionally' made the same thing, is a key point of the Aristotle-Thomas school of knowledge. The famous Latin saying that truth is *adequatio rei et intellectus* expresses this idea in brief terms.

Scholars of the theory of legal proof have underlined the need to use a concept of logic truth based on such an aphorism to fulfil the aim of the criminal proceedings, rejecting any other way of defining the concept of truth that may be derived from the corresponding philosophical schools, because they do not permit judges and juries to reach it.

The concept of truthfulness needs to make real sense, and it is only reached by using realistic metaphysics. Epistemology cannot be reduced to methodology, and methodology cannot be reduced to logic or sociology. As M. Artigas says (Artigas, 1992): "The value of scientific knowledge depends on a philosophical base which transcends the experimental methodology and its foundation."

From a moderate realist view it can be understood that evidence is the cause of certainty, and that certainty is only a state of mind, as uncertainty is as well. Truthfulness can only be a predicate of judgements, and only when they are made. Logic truth refers to reality comparing what is understood with what is known in reality. Therefore, any proper judgement is true or false, and there is no other possible intermediate state of affairs with respect to reality.

This philosophical way of understanding what the act of knowledge is allows us to be more precise on what the act of measuring uncertainty is. Some intellectual efforts to elucidate terms commonly used in conclusions of forensic reports have been carried out. The opinion stating that the definition of probability as a degree of belief in the occurrence of some event should be changed according to the degree of credibility in its occurrence has been defended. This change is not merely a cosmetic change of appearance because behind the concept of belief there is a way of understanding what knowledge or belief means.

It is believed that to identify subjective judgements with beliefs is not epistemologically right. In our view, it is a consequence of using the verbs 'to know' and 'to believe', or similarly, the terms 'knowledge' and 'belief', as if they had the same meaning.

It is proposed to use the wording 'degree of credibility' instead of 'degree of belief' to underline that subjective judgements are based on knowledge and faith. The former is related to personal knowledge, information and experience, while the latter is related to personal credibility in the truthfulness of something based on the knowledge, information and experience initially acquired by others in the chain of data transmission. In our view, this is quite a complete description of the meaning of subjective judgements.

It has been shown that probability is not commonly a direct method to measure uncertainty and, therefore, some mathematical transformations are needed to do it. According to the Shannon's information theory and the derived concept of normalised entropy, this concept could be used to measure uncertainty in a direct way and within the same numerical rank used for probabilities, that is, the closed interval [0,1], when two mutually exclusive and exhaustive propositions are alleged by the parties.

In addition, the widespread use in the scientific literature of the concept of abduction as a new way of reasoning needed for explaining the probabilistic inference process to evaluate evidence has been criticised. It is argued that it is a way of induction, specifically known as induction *in lato sensu*, characterised because although the premises of the syllogism are true, the conclusion will only be conjecture.

Some scientists support the existence of objective and subjective chances: the former in reality and the latter in mind. According to the philosophy of being, the objective chance is empty of being and the subjective one is empty of knowledge. Chance, in its relative sense, designs what is left without determination by a certain epistemic domain. Those who support the existence of ontological indeterminacy (objective chance) infringe upon the principle of non-contradiction. The 'other' of necessity is not chance but contingency.

5.2 Some legal conclusions

It is believed that any legal reform dealing with scientific evidence in court needs to preserve the legal and constitutional guarantees while aligning itself with state-of-the-art scientific knowledge. The very nature of forensic evidence reliability has been challenged in the last two decades, and a new paradigm is spreading steadfastly across countries towards a more solid and well-grounded forensic science. However, the 2002 Spanish Code of Criminal Procedure reform veered towards the opposite direction: objectivity and reliability were attributed to expert evidence provided by official laboratories by way of the legal status of documentary evidence, instead of that of classical expert evidence and to official reports dealing with weight, purity and nature estimations on seized drugs, a status that could be expanded by case law or other legal reforms to other criminalistic areas. It is believed that the objectives of reform regarding the efficiency and simplicity of laboratory procedures for reporting and testifying in court can and must be met by satisfying simultaneously both legal guarantees and international scientific standards.

Furthermore, any supplementary guidelines to the ISO 17.025 standard on accrediting identity criminalistic tests, or any policy to improve this type of forensic service, should be based on a solid logical framework, appropriate to the efficient and scientific evaluation of evidence, as is

the case in a DNA comparison. If not, these measures could bring discredit to the accreditation system in forensic science since transposed conditional fallacies could not only be made but accredited!

This research provides reasoning to justify the assistance of forensic experts to judges and courts to assign prior odds when the evaluation of scientific evidence has been made by likelihood ratios. We consider it very relevant that the judicial reasoning used to defend a result related to a criminalistic comparison is based on logic laws, which are used in science and not in hunches.

Legal implications of the proposed expert assistance to judges and courts in a Continental judicial system such as the Spanish one have also been discussed. In accordance with the applicable Spanish law and jurisprudence, we believe that there are grounds to assist them in this way. On the one hand, this technical support could be thought of as an epistemological guarantee to preserve a rational conception of the assessment of scientific evidence. On the other hand, this kind of forensic practice could reinforce the principle of equality of arms in the legal procedure.

5.3 Some scientific conclusions

A real casework given as an example for applying the new recommendations for making evaluative conclusions coming from ENFSI in forensic speaker recognition is shown in detail. This example has been included, albeit summarily, in the ENFSI Guidelines for Evaluative Reporting in Forensic Science financed by ISEC funds as a Monopoly project.

The mathematical tool proposed for the assistance of forensic experts to judges and courts to assign prior odds when the evaluation of scientific evidence has been made by likelihood ratios is a Bayesian network, because it is a logical structure of reasoning under uncertainty based mainly on Bayes' Theorem and Jeffrey's rule. These rules ensure that the inferential process strictly respects the probability laws. No one can reasonably say that the probabilistic assignments that judges and courts have come to are arbitrary, unfounded, unscientific, intentionally skewed or random.

The prior odds that have to be assigned in a Bayesian inference framework related to an expert report can be assigned numerically in line with generalisations and common sense in many cases. For discussion about the variation of priors on posteriors in paternity cases (see Ellman et al., 1979). A real case (voice comparison) has been discussed in detail in this thesis to illustrate how to implement in practice the inferential assistance suggested.

Future practical research will be needed to make good use of the possibilities that Bayesian networks can offer to judges and courts (and the parties to the proceedings) to make inferences, not only to assess coherently the evidence but also to make them during the procedural stages. This work is only the first of such possibilities (see Taroni et al., 2014, for further details).

This research also contributes a novel tool for detecting likelihood ratios presenting misleading evidence that is too high, with the proposal of so-called limit Tippett plots. Based on previous contributions (Royall, 1997; Royall, 2000, pp. 760–780; Aitken and Taroni, 2004), the behaviour of the bounds of the probability of observing strong misleading evidence in a set of LR values has been described. Two main conditions have been examined: unknown data distributions

(universal bounds) and normal data distributions. Thus, Tippett plots, including the theoretical bounds of the probability of misleading evidence, namely limit Tippett plots, have been proposed for detecting anomalous behaviour in LR values. This proposal is useful for examiners conducting experiments with the aim of measuring the performance of a given methodology using LRs.

PART III: PUBLISHED PAPERS

Chapter 6. THE AMERICAN 2009-NAS REPORT FROM THE VIEWPOINT OF SPAIN[7]

Abstract

In 2002, the Spanish Criminal Procedure Law was amended by converting expert reports for the determination of nature, amount and purity of drugs from official laboratories into documentary evidence, the objectivity and reliability of which are assumed by default. It has advantages both for the prosecution as well as the experts as they avoid testifying in Court in most cases. It could likely expand in Spain to other forensic fields, being the legislature who establishes what types of evidence are objective and reliable. However, Forensic Science is under a major redefinition. Error awareness in real cases (even in the ever-trusted domain of fingerprints), lack of solid scientific framework in different types of scientific evidence and fallacious reporting of results are highly topical questions. To strengthen Forensic Science in Spain, some good experiences are shown and the likelihood paradigm is proposed as the best solution to evaluate scientific evidence in Court.

Keywords: forensic science; evaluation of evidence; documentary evidence; likelihood paradigm; individualization paradigm; accreditation; objectivity; transparency; testability.

6.1 Multidisciplinary analysis of the impact of the NAS Report in Spain

The so-called traditional fields of expertise in criminalistics such as latent print comparison, firearms and tool mark comparison and handwriting examination, which have played an important role in many relevant criminal cases in all countries, have spread the use of categorical concluding statements on identity in expert reports. The untested claim that a forensic practitioner can link an unknown mark to a unique source, a faulty probabilistic intuition equating infrequency with uniqueness (Saks et al., 2008, pp. 199-219), has been a key determinant to achieve convictions, as prosecutors and Law Enforcement Agencies have relied on reporting results including categorical statements about identity to base their theses.

In Spain, the Supreme Court considers it sufficient to establish the identity of a person the comparison of latent prints with known fingerprints when forensic experts report 8 or 10 common minutiae (the disjunction 'or' is literal), e.g., equal topography (minutiae set in the same places) and morphology (types of minutiae), and ridge numbers between minutiae. Besides, there should be no natural dissimilarity between the latent print and fingerprint being compared (Spanish Supreme Court (henceforth, SSC) case law, e.g., see sentences: 15/06/1988, Ar. (Ar. stands for Aranzadi) 5024; 04/07/1988, Ar. 6477; 25/11/89, Ar. 9319; 04/07/1990, Ar. 6220; 15/03/1991, Ar. 2156; 02/12/1992, Ar. 9903; 02/11/1994, Ar. 8382; 04/11/94, Ar. 8563; 18/09/1995, Ar. 6379). In the case of several latent print-fingerprint comparisons from the same person, the Supreme Court acknowledges a greater degree of certainty on identity (SSC case law, e.g., see sentence: 25/11/1989, Ar. 9314). This was a consequence of the way Spanish forensic experts had defended fingerprinting evidence in Court in years past, which was considered highly effective by crime investigators. But it is a paradox that such a ruling, which determines the limits of reasonable doubt in fingerprint

[7] The full reference of the published paper is the following: LUCENA MOLINA JJ, PARDO IRANZO V, GONZALEZ RODRIGUEZ J, "Weakening Forensic Science in Spain: From Expert Evidence to Documentary Evidence". Journal of Forensic Sciences, 57(4), July 2012, pp. 952-963.

identification, could be used to wrongfully involve innocent people (Stacey et al., 2004, pp. 706-715). Individualization is still the key task of many forensic experts in the majority of countries worldwide.

Many scientists and commentators have refuted the individualization paradigm for decades from philosophical, logical and statistical standpoints (Saks et al., 2008, pp. 199-219; Stoney, 1991, pp. 197-199; Jeffreys, 1931; Royall, 1997; Jaynes, 2003; Aitken et al., 2004; Saks et al., 2005; Kaye, 2009). As E.T. Jaynes wrote in (Jaynes, 2003) a quote which he attributes to H. Jeffreys (Jeffreys, 1931): "The role of induction is not to tell us which predictions are right, but which predictions are indicated by our present knowledge … Moreover, it is only when the inductive inferences are wrong that new things are learned about the real world." It is important in forensic identification to be aware of the role of induction in science and, consequently, the role of statistics.

On the other hand, forensic identification fallacies, such as the often published prosecution fallacy, continue to be either unknown or misunderstood by some forensic experts, prosecutors and judges all over the world (Aitken et al., 2004). Very different reactions have been noted, especially among European countries and their laboratories (Norgaard et al., 2009), from the time scientists and statisticians began explaining and publishing those fallacies in scientific journals and books. For example, after years of internal debate, in 2009 the Swedish National Laboratory of Forensic Science (SKL) announced its decision to unify the manner conclusions are drawn in evaluative comparisons, regardless of the fields of expertise involved, in order to provide the same kind of information to courts whatever the type of evidence analyzed. However, many other laboratories have barely reacted or are awaiting superior orders, as is the case of Spain.

There shall be a focus on the role of science in the evaluation of evidence by forensic experts and its implications on Spanish law. Other important roles of forensic science such as investigative or intelligence tools are omitted. Questions such as: "What does scientific methodology consist of?"; "What methodological requirements should be asked in any forensic report?"; "What does objectivity mean in science (Agazzi, 1979, p. 121; Berger et al., 1988, pp. 159-165; Evett, 1996, pp. 118-122)?"; "How should data obtained from experiments or observations be interpreted as evidence (Royall, 1997)?"; "What is the role of forensic experts evaluating evidence (Aitken et al., 2004.)?", as well as many others, do not have simple answers. To quote from (National Research Council, 2009), in the Reporting Results Section, it is absolutely necessary "to provide the essential building blocks for the proper assessment and communication of forensic findings." There will be some mention of how the accreditation and suitable statistical interpretation of scientific evidence are pillars to guarantee the reliability requested by Courts and crime investigators in reporting results.

Differences between the judicial and scientific contexts with respect to essential aspects of science applied to solving criminal cases will be shown. The massive, ever increasing number of requests for expert reports received by official laboratories and the chronic lack of human and material resources have led to the following measures designed to speed up criminal proceedings: defense of forensic reports by videoconference; the presence of only one expert before the Court in certain types of proceedings and also, legal reforms in Spain. An example of the latter is the addition in 2002 of a second paragraph to Article (Art. from now on) 788 of the Spanish Code of Criminal Procedure (henceforth, LECrim) converting expert evidence on the determination of the nature, amount and purity of the drugs seized into documentary evidence (Pardo, 2008). Since 2002, based on foregoing legal amendment, some Sentences

passed by different Spanish Courts applied the analogy of drug examination, that is, the consideration as documentary evidence, to other forensic fields. Consequences of such judicial change are relevant in evidence interpretation as well as in the work of Courts, prosecutors and lawyers. Needless to say, this reform was carried out without previous scientific debate.

Whether by means of case law or legal reform, the many years of work by Spanish official forensic experts in fingerprint comparisons or drug examinations are finally acknowledged to be "scientifically reliable." Whereas on the one hand, for instance, in the United States (National Research Council, 2009) or United Kingdom (The Law Commission Consultation Paper No 190, 2011) technical committees are appointed to provide professional advisory documents to law commissions before discussing bills related to giving expert evidence in Courts in Congress or Parliament, on the other hand, in Spain, the priority is given to the official forensic expert's independence. The official character of forensic experts and the fact they belong to Public Institutions afford the Courts a guarantee of scientific competence; for this reason, controversies such as those that recently happened in the United States are simply unthinkable.

Undoubtedly, it is relevant to determine the causes of the controversy from (Kennedy, 2003, p. 1625; www.innocenceproject.org; Dror, 2006, pp: 600-610; National Academy of Sciences, 1979) and to contribute proposals about how efficient procedural justice can be made compatible - without changing the true nature of evidence in criminal proceedings - with a rigorous scientific method used in presenting conclusions in forensic reports. In Europe, the history of the so-called Voiceprint methodology, e.g., to recognize people by voice, could be reproduced again in many forensic fields in the absence of a previous controversy. The National Academy of Sciences of the United States published a technical document in February 1979 (National Academy of Sciences, 1979) previously requested by the FBI which set out to analyze the scientific bases of such a methodology. Once published, the FBI decided against its use in making categorical identity statements. Nonetheless, other voiceprint experts in many other countries continued to use such statements. In other words, the opportunity to strengthen Forensic Science in the United States and beyond could be lost if the need to establish the fundamentals of the scientific method as applied to forensic practice is ignored.

The structure of this paper is as follows: firstly, an overview is provided on criminalistic identification to help put the Spanish case in a global context, establishing what is happening nowadays in forensics all over the world; secondly, the role of science in criminalistic identification pointing out the discussion in these three words: objectivity, transparency and testability; thirdly, a detailed analysis of the Spanish legal reform wherein expert evidence is transformed into documentary evidence, together with the implications for the interpretation of evidence in Courts by experts; fourthly and lastly, some opinions and possible solutions that would enable compatibility between Spanish legal reform objectives and the specific demands of science.

6.1.1 An overview of criminalistic identification

The criminalistic environment is extremely complex. It involves both private and public laboratories, with the latter including Public Institutions belonging to ministries or governmental departments of Interior, Justice, Health, Customs, Commerce, Treasury, Defense, and others (i.e., Law Enforcement Agencies, Secret Services, Prosecutor Offices, Forensic Medicine or Toxicology Institutes, amongst other possible ones). The so-called 'Fragmented system' in the United States in (National Research Council, 2009), is not far away

from what is happening in the rest of the world. In Spain there are five laboratories that belong to the European Network of Forensic Science Institutes (henceforth, ENFSI), two of them are dependent on the Ministry of Interior (National Police and Guardia Civil), two on Regional Governments (Ertzaintza in the Basque Country, and Mossos d'Esquadra in Catalonia), and another one, on the Ministry of Justice (Toxicology and Forensic Sciences). There is hardly any technical coordination between them on forensic report interpretation, and even different departments in each laboratory report their findings differently.

However, while commentators were already warning in the eighties and nineties (Diaconis, 1981, pp. 333-334; Thompson et al., 1987, pp. 167-187; Saks et al., 1991, pp. 361-372; Fienberg et al., 1991, pp. 265-270) about the need to avoid 'fallacies' or pitfalls when forensic experts assessed analytical results, nowadays these 'fallacies' are ignored, more or less consciously, in quite a few laboratories of many countries (Aitken et al., 2004; National Research Council, 2009). The individualization paradigm (Saks et al., 2008) continues to be widely spread, and categorical statements in reporting results are supported by many forensic experts and laboratories.

In Europe, ENFSI is undertaking its own self-transformation within each Working Group (WG). Despite being aware of this scientific problem since the very first conference of the European Academy of Forensic Sciences (1997 - Switzerland), only a few of those WG's have reached a consensus on how to interpret evidence in Court. As stated in Part 2, R. & Clark (The Law Commission Consultation Paper No 190, 2011), "those experts are trying to solve problems outside their fields of expertise." This explains why the ENFSI Board strongly supports initiatives such as the on-line training courses on statistics applied to forensic fields; research on the interpretation of evidence or the successful FORSTAT Workshops (see FORSTAT at Bibliography) aimed at improving the statistical background of the European forensic experts.

On 4 June 2010, the current Chairman of the ENFSI Board sent a letter to the Head of the Criminal Law Division of the Directorate General of Human Rights & Legal Affairs of the Council of Europe. The letter endorsed the report titled "Scientific evidence in Europe – Admissibility, Appraisal and Equality of Arms", written by Christophe Champod and Joëlle Vuille (Champod et al., 2010), both of the University of Lausanne (Switzerland). It underlined, on behalf of the ENFSI Board, the parallelism between the strategic objectives of ENFSI and the recommendations given in that document, specifically "the importance of the scientific interpretation of results."

The forensic scientific community can think itself lucky with the appearance of the Standards for the Formulation of Evaluative Forensic Science Expert Opinion (AFSP ,2009, pp. 161-164) in the aforementioned document, as a starting point towards a solid logical framework in the evaluation of evidence in Europe. Its scope is defined as "Forensic Expert Opinion formulated in the Evaluative or Evidential mode across all scientific disciplines." Probabilistic inference in forensic science has achieved enough scientific maturity due to a number of well-considered proposals dealing with terminology, the logical framework applicable and the role of the forensic expert in evidence evaluation. The aforementioned Standards are just one such example. Also, Part 1 (The Law Commission Consultation Paper No 190, 2011) states that "scientific knowledge is continuously advanced as more empirical research is undertaken, so it is inevitable that some hypotheses will come to be modified", and "special caution is also needed where expert opinion evidence is not just relied upon as additional material to support a prosecution but is fundamental to it" and it is thus worth being aware of the growing

concern in the scientific and judicial communities about how scientific evidence is being tendered in Courts.

It should be acknowledged that categorical statements have usually been welcomed by the legal system. In Spain, handwriting examinations have been used to achieve convictions in terrorism cases. For instance, in bomb-production training courses for terrorists, it is frequent for those attending to take their notes by hand. When Law Enforcement Agencies arrested suspects and handwritten notebooks seized, categorical statements of authorship clearly helped to achieve convictions for belonging to a terrorist organization, resulting in a minimum imprisonment of six years and one day. The strength of the evidence assessed in terms of categorical statements by forensic experts has always had extraordinary importance for the efficacy of any Law Enforcement and Justice system, making it difficult to fight against this misunderstanding. Self-evidently, fingerprint examinations and DNA analyses have played a similar role in terrorism and other criminal organization investigations. But while DNA procedure was tested in depth from a scientific point of view from its inception, fingerprinting has not been tested to a similar extent (National Research Council, 2009; Bohan, 2010, pp. 5-7; Spinney, 2010, pp. 344-346). Nevertheless, categorical statements on identity have been defended by fingerprint experts all around the world for decades. Undoubtedly, the so-called DNA paradigm marks a new era in the assessment of evidence in Courts (National Research Council, 2009; Gonzalez et al., 2007, pp. 2104-2115).

It is believed that it is not possible to avoid a strong comparison between the traditional criminalistic context and the likelihood paradigm (Royall, 1997) from the scientific community. Terminology and logical framework problems related to the evaluation of evidence by forensic scientists have been discussion points within laboratories as is clearly addressed in Chapter 6: Improving Methods, Practice, and Performance in Forensic Science, Section: Reporting Results (National Research Council, 2009): “There is a critical need in most fields of forensic science to raise the standards for reporting and testifying about the results of investigations. For example, many terms are used by forensic experts in reports and in court testimony to describe findings, conclusions, and the degrees of association between evidentiary material (e.g., hairs, fingerprints, fibers) and particular people or objects. Such terms include but are not limited to ‘match’, ‘consistent with’, ‘identical’, ‘similar in all respects tested’, and ‘cannot be excluded as the source of’. The use of such terms can have a profound effect on how the Trier of fact in a criminal or civil matter perceives and evaluates evidence. Yet the forensic science disciplines have not reached agreement or consensus on the precise meaning of any of these terms. Although some disciplines have developed vocabulary and scales to be used in reporting results, they have not become standard practice. This imprecision in vocabulary stems in part from the paucity of research in forensic science and the corresponding limitations in interpreting the results of forensic analyses.”

This comparison explains to a certain extent a controversial recommendation recently given in the United States (National Research Council, 2009), the already famous Recommendation 4: “To improve the scientific bases of forensic science examinations and to maximize independence from or autonomy within the law enforcement community, Congress should authorize and appropriate incentive funds … for the purpose of removing all public forensic laboratories and facilities from the administrative control of law enforcement agencies or prosecutors’ offices”, suggesting independence from or autonomy within law enforcement agencies. It is believed that appropriate changes should be taken in official labs regarding quality assurance procedures and the full integration of the likelihood paradigm in their daily reasoning and reporting work.

6.1.2 The role of Science in Criminalistics

6.1.2.1 What can science actually do?

As R. Royall wrote (Royall, 1997) in the Preface: “Science looks for statistics for help in interpreting data. Statistics is assumed to provide objective methods for representing scientific data as evidence and for measuring the strength of that evidence. Statistics serves science in other ways as well … But its most important task is to provide objective quantitative alternatives to personal opinions for interpreting the evidence produced by experiments and observational studies. In this role statistics has made fundamental contributions to science”, therefore the most important task of science regarding reporting results is to provide objective methods for evaluating evidence after carrying out the relevant analyses according to accredited procedures.

Problems need to be acknowledged along the last century as regards interpreting data as evidence by statisticians. The Neyman-Pearson-Wald decision-making and the Fisherian (p-value) prevailing paradigms for the last century were erroneously applied as solutions to interpret data as evidence (Royall, 1997): “… All is not well, however. Standard statistical methods regularly led to the misinterpretation of results of scientific studies … These misinterpretations were not a consequence of scientists misusing statistics. They reflect instead a chronic defect in current theories of statistics. These problems exist because the discipline of statistics has neglected a key question for which it is responsible: when does a given set of observations support one hypothesis over another? In other words, when is it right to say that the observations are evidence in favor of one hypothesis vis-à-vis another?”

Royall (Royall, 1997) argues that: “The concept of evidence is missing altogether in the Neyman-Pearson theory – its author insisted that rejecting one statistical hypothesis in favor of another signifies a decision to act in a certain way, and nothing else.” Though significance tests (p-value procedures and rejection trials) have attempted to do what Neyman-Pearson does not (to quantify the strength of statistical evidence), Royall (Royall, 1997) shows that significance tests fail in this endeavor because they rest on the faulty foundation of the law of improbability.

As the same author concludes (Royall, 1997): “The likelihood paradigm represents a solution to the dilemma that science has faced since the emergence of the modern Bayesian movement in statistics in the 1950s”; in other words, the logical defects and inconsistencies of frequentist statistical methods (of both the Neyman-Pearson and the Fisher school).

It has oftentimes been reiterated that categorical statements on identity are found in forensic reports all over the world as set out in Chapter 1: Introduction, Section: Pressures on the Forensic Science System, Paragraph: Questionable or Questioned Science (National Research Council, 2009) the following: “Some forensic science methods have as their goal the “individualization” of specific types of evidence (typically shoe and tire impressions, dermal ridge prints, tool marks and firearms and handwriting). Analysts using such methods believe that unique markings are acquired by a source item in random fashion and that such uniqueness is faithfully transmitted from the source item to the evidence item being examined (or in the case of handwriting, that individuals acquire habits which result in unique handwriting). When the evidence and putative source items are compared, a conclusion of individualization implies that the evidence originated from that source, to the exclusion of all other possible sources. The determination of uniqueness requires measurements of object

attributes, data collected on the population frequency of variation in these attributes, testing of attribute independence, and calculations of the probability that different objects share a common set of observable attributes. It is very important for the results of the investigation to be made public so that they can be reviewed, checked by others, criticized, and then revised, and this has not been done for some of the forensic science disciplines."

Even though in the last two decades forensic statisticians and scientists have emphasized the importance of avoiding fallacies in reporting conclusions and in following the likelihood paradigm (Aitken et al., 2004), and some international forensic institutions such as ENFSI have made efforts in both directions, nowadays there are still official laboratories in Europe, even members of ENFSI, using categorical frameworks to interpret data as evidence, mainly in so-called traditional criminalistic fields of expertise such as fingerprints, firearms and tool mark comparison or handwriting examination.

The individualization paradigm has been harshly criticized in (Champod, 2009), where the author emphasized that this paradigm should always be avoided in the forensic science community as it is not necessary in order to be useful, not even in DNA. It is a very crucial question to understand what philosophers of science, experts in logic and probability theory, and forensic statisticians, among others scientists, have published in recent decades about how to make an inference to evaluate the evidence, in the awareness that "individualization" implies an unjustified leap of faith (Stoney, 1991, p. 197-199). As Kaye D.H. comments (Kaye, 2009) in footnote number 39: "Accepting any inference about any population parameter is never risk-free. 'The gap between the sample and the population will always require a leap of faith.' (Good, 2003) The only issue worth debating is the length of the leap."

Once clarified that the interpretation of data as evidence is the main task of forensic experts in trials, how should evidence be evaluated in practice? The so-called likelihood paradigm by Royall (Royall, 1997) solves the problem, and as an example, the above mentioned Standards for the Formulation of Evaluative Forensic Science Expert Opinion (AFSP, 2009, pp. 161-164) from the AFSP does what Royall recommends. In essence, what it is necessary to distinguish is the difference between the following questions once observations have been made: "What is to be believed?", "What is to be done?", and "What does that observation say about A versus B, when A and B are, generally speaking, simple hypotheses?", that is, "How should this observation be interpreted as evidence regarding A versus B? (Royall, 1997)?"

6.1.2.2 Objectivity

Experimental science claims to be objective knowledge. Objectivity has a double meaning: inter-subjective validity on the one hand and something belonging to the real world on the other hand. The former suggests that scientific methods and results are valid for any person, whatever their personal convictions. Both meanings are related to each other.

Scientific inter-subjectivity cannot be interpreted as perfect logical demonstration: scientific demonstrations cannot be formulated using only logical links. Epistemologists cite contextual demonstrations, emphasizing their dependency on the context of the objectivation adopted. The concept of objectivation is used here in the same sense as Evandro Agazzi does. For example, in (Agazzi, 1979, p. 121): "… if we agree to investigate reality using a precision scale, a chronometer, or a rigid rule to verify propositions with predicates as 'mass', 'time' or 'length' according to standardized measurement procedures, then we can establish the immediate truthfulness or falseness of a number of base-statements and will formulate empirical

assertions of classical mechanics" (translation carried out by the authors). As Artigas M. (Artigas, 1992) defines it, objectivation is a set of theoretical and practical procedures used to make scientific objects, and asserts that objectivation is the key to establish inter-subjectivity in science.

However, objectivity, when understood as lack of personal influence in forensic findings is illusory (Berger et al., 1988, pp. 159-165; Evett, 1996, pp. 118-122). The belief in the ability of statistics to inject and guarantee objectivity takes root in the already mentioned paradigms predominant in the last century. In this sense, difficulties can be predicted for those who want to understand reporting results given by forensic experts in non-intuitive disciplines. "Intuitive" forensic science disciplines are defined as those whose results are supported by descriptions or audiovisual means which facilitate educated laymen's understanding of scientific reasons expounded by scientists in their reports.

Unlike traditional criminalistics, the newer disciplines address complex phenomena whose results cannot be easily interpreted by laymen. Even if results from intuitive disciplines were to be easily understood by everybody, the inference process in evidence evaluation from data, as has been mentioned earlier repeatedly, is not a trivial matter as it can be full of misunderstandings. Therefore, the apparent objective character of a forensic report does not imply that it could be easily understood by anyone in the same way, nor that it is actually objective. The problem of interpretation will be always present, even in the so-called factual reports (AFSP, 2009, pp. 161-164).

Objectivity from an accreditation viewpoint has to do with validated methodologies, calibrated and verified equipment, detection and quantification thresholds, certified reference material, and uncertainty, accuracy and precision estimation when measuring. It is also related to the assurance of repeatability and reproducibility of results; selectivity, specificity, and traceability of procedures; to proficiency testing and to be subjected to internal and external audits by accredited experts and national organizations for test accreditation. Some scientists have summarized all those requirements in two words: transparency and testability (Gonzalez, 2007, pp. 2104-2115).

Unfortunately, this kind of objectivity proved to be compatible with mistaken ways of evaluating evidence (using verbal scales of probability of propositions claimed by parties or even categorical conclusions).

ISO 17.025 accreditation guarantees the technical competence of laboratories and the reliable results of tests and calibrations. The Spanish National Entity for Accreditation (henceforth, ENAC) is the Spanish entity responsible for establishing the accreditation system following international rules and EU policy in this matter.

There are no supplementary guidelines in the EU similar to those of the American Society of Crime Laboratory Directors /Laboratory Accreditation Board (ASCLD/LAB) in the US. So far there has been an initiative promoted by ENFSI based on the experience of more advanced European laboratories in this field so as to help others to accredit their official tests (known as the EMFA project - European Mentorship for Forensic Accreditation Project), and EU countries will have to comply with the provisions of the European Commission Directive on DNA and fingerprint analyses accreditations.

Several months ago, ENAC promoted ad hoc working groups before auditing, for the first time, new specific criminalistic tests, carried out almost exclusively or preferably by official forensic laboratories (fingerprinting and handwriting examinations in particular). The meetings were attended by practitioners of the aforementioned laboratories - carefully chosen by their directors - and members of ENAC. The practitioners showed ENFSI and Interpol's documents (Interpol European Expert Group on Fingerprint Identification, 2011), and some forensic bibliography on which to base their tests, but difficulties arose when they asked ENAC members to accredit their conclusions: ENAC does not allow the interpretation of results to be accredited.

In the evaluation of evidence it is very important to distinguish suitably between results (data), the interpretation of results (interpretation of data as evidence) and the conclusions (what experts answer to questions made by petitioners of forensic reports). In fingerprint and handwriting comparisons inductive inference should be made once data is obtained and hence the conclusions are clearly interpretations (evaluation of evidence). The so-called evaluative opinions (AFSP, 2009, pp. 161-164) are not logically compatible with categorical or probabilistic statements about propositions. However, there are already quite a few examples in Europe regarding accreditations under ISO 17.025 with accredited conclusions made using, for instance, categorical or probabilistic statements about propositions, like in the fields of handwriting or fingerprint comparisons. Such was the case in Spain in two main laboratories belonging to ENFSI.

In this context, it is believed that results may be accredited, as well as the interpretation of results within a solid logical framework such as the likelihood paradigm (not at all in the individualization paradigm), and the conclusions as well, since they may be composed for any or both of them, i.e., results and interpretation of results.

Does it make sense to plan for accreditation deadlines in respect of classification pattern problems which have not yet been established on solid logical bases and in the context of the evaluation of evidence? Such a policy could result in the general discrediting of the accreditation system used in forensic science. It is believed that this could be the case with respect to fingerprint comparisons in Europe due to the current policy of the European Commission on this matter (Council framework decision 2009/905/JHA, 2009, pp. 14-16).

6.1.2.3 Scientific reliability in the Spanish judicial system

There are two different systems in Spain in the evaluation of admitted evidence. The first one is called the "Free Evaluation System" and the second is known as the "Legal Evaluation System." If the rationale of the latter is to consider that this system implies a higher legal certainty, the former system is said to provide greater rationality.

In the "Legal Evaluation System", the statutory law indicates to the judge the value of evidence without taking into consideration the judge's opinion. For example, in Spanish civil procedure the evaluation of public documents is carried out according to this system (Art. 319 of the Civil Procedure Law) and therefore if a public document is not challenged the judge must take it into account as true even if he were to have any doubt about it.

On the other hand, in the "Free Evaluation System", the judge evaluates evidence freely. This system applies in Spanish criminal procedure to all types of evidence. This view notwithstanding, it is also true that in practice, judges, who obviously are not scientists,

consider that documents coming from official laboratories, if not challenged, have a privileged strength from an evidence standpoint. Therefore, in trial, it is very important for the expert witness to interpret or explain the expert's report data ably, so as to enable the judge to make a proper evaluation thereof.

Taking into account the four options in evaluating evidentiary reliability in the common Law as quoted in Part 4, Introduction, 4.3 (The Law Commission Consultation Paper No 190, 2011): (1) Exclusionary discretion without guidance; (2) Exclusionary discretion with guidance; (3) An admissibility rule requiring consensus amongst experts in the field; (4) An admissibility rule requiring the trial judge to assess the evidentiary reliability of the tendered evidence, and focusing exclusively on the issue at hand about scientific reliability, the principle of free evaluation of evidence in the Spanish judicial system would fit into the first option. Therefore, both the pros and cons about the scientific reliability of expert evidence can be found in case laws, but the above-mentioned legal reform in 2002 of the LECrim provided a new way for considering some forensic reports (determination of nature, amount and purity of drugs) as scientifically reliable. In this case, scientific reliability has been legally established, and there have already been case law extending this scientific reliability condition to other types of forensic reports.

Since Spanish legislature is responsible for establishing which forensic reports should be deemed scientifically reliable, such a qualification derives solely from the official laboratories (these are the main providers of forensic services within a legal system based on the continental law), actual sources of arguments used by Courts in Case law. Therefore, the key to considering which forensic reports are reliable and which others are not is precisely the opinion held by said official laboratories. This was exactly what happened at the Supreme Court in Spain as regards fingerprint examinations, but in that case through their case law.

6.1.3 The legal and scientific status of forensic evidence in Spain

6.1.3.1 Spanish model for incriminating evidence (Guerrero Zaplana, 1999)

Art. 741 of the LECrim establishes an incriminating evidence model according to which only the actions taken in accordance with the principles of orality (right of the accused to be heard, oral testimony of the witnesses, oral deliberations etc.), contradiction (cross-examination), immediacy (presence of the parties and their witnesses at the trial) and publicity (public trial) can be referred to as such. According to constitutional case law, the only means of proof (see http://ec.europa.eu/civiljustice/evidence/evidence_spa_es.htm, to understand the concept "means of proof" used in the Spanish law [accessed January 12, 2017]) that can be used to invalidate the presumption of innocence are those used in trial, in addition to evidence existing before trial, i.e., impossible or difficult to produce at the trial.

However, this idea should not be understood radically which leads to the denial of the value as evidence of police and legal inquiry proceedings even if conducted in accordance with Spanish constitutional and procedural laws; but rather, it is required - for the recognition of this efficacy of evidence existing before trial – that they should be reproduced at trial under conditions that give the defense lawyer the opportunity to contest them or to tacitly accept their content (SSC case law, e.g., see sentence dated April 27, 1998).

Exceptions to this rule are the assumptions of advance evidence and evidence existing before trial as long as defense and contradiction rights are granted. However, when reports are drawn

up by the officially appointed experts, in collegiate form, and enjoying the permanence and job tenure of the civil servant, usually distanced from the case in point, with high levels of specialization and assigned to bodies endowed with the costly resources which modern analysis techniques require, it does not seem out of place to regard them as objective, impartial and independent, which in principle gives them value as evidence without procedural contradiction. However, their reports can be contested, either by asking in writing for extensions or explanations so that they can be accepted at the trial as documentary evidence, or by demanding the experts' presence during the proceedings.

This evidence is regarded as evidence existing before trial because of the nature of the procedural system, since it is not possible for the experts working for the above mentioned official institutions, departments or laboratories cannot carry out their tasks if they have to always be present at the trial to ratify their reports.

It is not disregarded that the prosecutors have to provide the evidence and this has to be carried out in court in contradictory fashion and no procedural responsibility in this regard. However, in view of the guarantees offered by said export reports, they are considered as evidence - in theory – without prejudice to the right of the accused to challenge its impartiality (by way of the possibility of refusal, which is always open) and to submit them to questioning if it is so entitled, thereby being guaranteed the right to defense (SSC case law, e.g., see sentences dated November 11, 1993, May 21, 1997 and June 24, 1998).

To understand how evidence is considered in the LECrim a brief explanation of how the Spanish criminal proceeding is made up is in order. There are two distinct, different stages: the first one is called investigative stage and the second one, trial stage. The former is mainly addressed to finding out circumstances around a crime and its authorship. During this phase investigative actions are undertaken (for example, scene examination or questioning), but these activities do not actually constitute legal evidence and therefore they cannot invalidate the presumption of innocence, nor be enough to convict. On the other hand, they can be used for the investigating magistrate to decide to shelve the record of proceedings or start the second stage ("the trial"). This phase, today and except in the case of penal proceedings involving minors, falls within the responsibility of an investigating judge though nowadays there are proposals asking for this responsibility to be assigned to the prosecutor in Spain. At the second stage, under a different Judge, the prosecutor makes charges and the evidence is provided - according to the four above mentioned principles (i.e., orality, immediacy, contradiction and publicity) and, as a general rule, the judge shall proffer a sentence convicting or acquitting based on the evidence given before the court.

Chapter III of Title III (called "the staging of the trial") of the LECrim establishes which evidence may be used in criminal proceedings and how it has to be given before the court in order to be taken into account when passing sentence. Said chapter distinguishes between the deposition by the accused (Art. 688 and ff.), witnesses (Art. 701 and ff.), expert evidence (Art. 723 and ff.), and documentary evidence (Art. 726). As a general rule, for the judge to consider the first three means of evidence, this requires the presence of people before the court, for example, an eyewitness who may state what he saw or heard or a forensic expert explaining in court the conclusions of the report issued by his laboratory. However, when the evidence is documentary, there is no declaration in court but rather magistrates can read the document for themselves.

As regards the distinction between expert evidence and documentary evidence in Spanish law, the former is required when scientific or artistic knowledge is needed to assess some fact which the judge does not have and must thus ask an expert to convey this to him. For example, in an analysis about narcotic substances, the judge – who is not a scientist - does not know whether the seized drugs is cocaine, heroin or just flour and he thus requires the aid of an expert who tells him, "What substance it is" and "What its weight and purity are." This report, carried out by an expert, is expert evidence. On the other hand, documentary evidence is (usually) a document, but in which there is no scientific analysis. For example, a letter sent by someone to someone else telling him that on a trip to a certain country he acquired a certain amount of hashish or, to use another example, a bank statement.

It should be borne in mind that expert reports are not a means of proof like others: it does not adduce facts to the process like witness and documentary evidence and depositions, as established by the LECrim; however, it is used by the judge as an auxiliary means to duly interpret certain facts that are already part of the process, provided through other means of proof. Moreover, it is a necessary element, since the judge does not have the technical knowledge necessary to solve the problem.

The fact that an expert is needed does not mean that his report is binding on the judge: firstly, it would be impossible to solve a case when reports are contradictory (a clear paradox of categorical statement reporting is that contradictory categorical statements are possible to be found in Court, which should be impossible if the disciplines were, as claimed, scientific). According to case law, if there are several expert reports on the same issue, it is the impartial organ, i.e., the Court presiding over the evidence, that decides which will prevail. Elements from different reports could also be used (SSC case law, e.g., see sentence dated March 7, 1987). Should there be only one expert report, it would not be binding on the judge because they are not by themselves incontrovertible truth (SSC case law, e.g., see sentence dated July 8, 1987).

The value as evidence of expert reports relies neither in the statements nor in the authors' position and standing, but in their scientific basis and grounds. The statements or conclusions of a greater rational quality will prevail, without ignoring other auxiliary criteria as coincident majority or their not being related to the interests of the parties (SSC case law, e.g., see sentence dated May 11, 1987).

In any trial, witness, documentary and expert evidence are usually examined; the results might be mutually contradictory, both within and between types. Expert reports as well as other means of proof are subjected to the principle of free appraisal of the proof which basically requires a joint assessment, without giving a priori more value to one of them. Thus, if apart from the expert report, different evidence has been examined in relation to a specific issue and with different results, the Court has the authority to make a joint appraisal of the evidence to ascertain that the truth of the facts to be clarified is not the truth set out in the expert report, but the truth offered by the other means of proof (SSC case law, e.g., see sentence dated October 28, 1998). Therefore, the expert does not adduce facts to the process, but opinions; neither does he/she aver on the existence or non-existence of facts.

Expert reports, i.e., opinions about certain facts, can be finally accepted or not by the court. According to case law, the judge dissents, upon reasoning, from the conclusions set in the expert reports, particularly when the report does not express a certainty, but merely a possibility, and an alternative possibility rather than an exclusion (SSC case law, e.g., see

sentence dated April 29th 1997). The court can only diverge from the conclusions drawn by the expert when there are objective reasons that allow or justify it. The reasons that have led the court to diverge from the expert reports should be clearly specified so as to avoid the risk or the suspicion of arbitrariness.

To sum up, expert evidence is not binding on the Judges. An expert report is a practical or scientific advice that affords a better insight of the reality underlying a problem, submitted to the consideration of the Judges; in accordance with case law, a judge may not yield or relinquish his/her evaluating responsibility (SSC case law, e.g., see sentence dated October 28, 1997).

6.1.3.2 A critical review of the Spanish legal reform to evaluate the nature, weight and purity of drugs in criminal proceedings

In 2002, on the occasion of the endorsement of the Organic Law 9/2002 of 10 December, which changed the Spanish Penal and Civil Codes on child abduction, the LECrim was amended by adding a second paragraph to Art. 788. According to this principle "the expert report can be presented by a single expert. Within this procedure, a report issued by an official laboratory in relation to the nature, amount and purity of drugs will be considered as documentary evidence as long as the fulfillment of the scientific protocols approved by the relevant standards has been specified." In other words, in the scope of the Spanish brief procedure certain expert reports are subjected to a legal consideration appropriate for documents, i.e., it transforms expert evidence into documentary evidence.

The legislature justifies this reform on the basis that expert reports prepared by official institutions regarding the weight, amount and quality of drugs are not, strictly speaking, scientific testing. Their repetition and execution according to specific scientific procedures (official protocols), they say, makes it necessary to consider them as documents of an objective reality and, it has to be added, it is said, as the impartiality and job tenure of the relevant officers in charge (SSC case law, e.g., see sentence dated February 3, 2009) as well. However, although this is not the only point, because of this reform some experts will not have to attend trial, which in turn will prevent the judge from deciding between interrupting the proceedings, in case of nonappearance, or rendering the evidence insufficient to nullify the presumption of innocence (Pardo, 2008; Diaz-Cabiale, 2003, p. 67).

It is commonly accepted practice by some jurists to distinguish between two different expert reports: those that they called "scientifically objective" and those based on opinions (Alvarado-Velloso, 2006). By means of testing (e.g., DNA tests), a "scientifically objective" report purports to verify the accuracy of a statement or to establish the certainty of some facts. An opinion report, on the other hand, does not intend to verify but rather, to assess or evaluate a fact or circumstances of a fact (e.g., a psychological report). Then, the legislature has decided that the analysis of drugs, which in its own opinion falls within the scope of "scientifically objective" reports, should not kept as such in the Spanish brief proceeding and rather, should be dealt with as documentary evidence. Since the above-mentioned amendment, the same legislature considers the above-quoted drug reports as documentary evidence in brief criminal proceedings while expert evidence in ordinary criminal proceedings: a clear inconsistency in the Spanish legislation!

Nevertheless, the legislature does not take into consideration that:

(i) According to the Art. 456 LECrim, the judge agrees with the production of an expert report whenever scientific or artistic knowledge is needed to learn of, or understand any fact or relevant circumstance. Conversely, a document intended for criminal proceedings is something different: it is an item (corpus) which shows or displays something (docuit) but additionally, it is "strange" to the process as the document should not be produced on the basis of the process (estraneità). In the case in point, scientific expertise is needed to determine the nature, amount and purity of the substances. That is the reason why the judge requires an expert (a person who is an outsider to the process and has the appropriate specific training) to produce an expert report (scientific testing). There is a clear reference here to expert evidence, not just documentary evidence.

(ii) Even if "scientifically objective" reports are presented, careful consideration is strongly advised for their interpretation (not only with regard to the conclusion on nature, amount and purity of the substances but also, on whether the official protocols have been fulfilled). The expert should transmit this interpretation to the judge, who is a lay person in this field (cf.: see Madrid Provincial Court case law num. 26/2004 of 29 April: the Public Prosecution proposes as documentary evidence the reading of page 104, where the techniques used to carry out the analysis are explained: "Color R., organic solvents extraction, CCT, EIA, GC-SM, HPLC. Besides this statement, it also includes an explanation of the terms expressing the richness of each product analyzed"). It is obvious that based only on this, a judge who is a lay person in this field will not be able to infer that the analyses were carried out in compliance with the scientific protocols approved by the relevant standards.

(iii) When testing is carried out there is also or there may be a variability regarding the conclusions. Suffice it to observe the test organized by ENFSI on the detection of the nature and purity of several drugs which is discussed in the Appendix.

It is believed that analyses on drugs are expert reports (forensic evidence), as are other types of scientific evidence (e.g., fingerprints, DNA, handwriting ...), and they may not be considered as simple documents (documentary evidence). On the other hand, as the likelihood paradigm is still not used in most official departments in official laboratories, there exists a wide variability in terms of conclusions. This variability could cause problems in the interpretation of conclusions. It is believed that such interpretation should be made by the expert and not by the judge himself by simply reading the report. In spite of this, the legislature has unrealistically deemed expert evidence to be documentary evidence. This is a relevant question as it is not only a change of designation, but it implies a change in the approach to the reports submitted during the process. In that respect, the following opinion is held:

(i) Since this reform was enacted, it is neither necessary to have the expert appointed by the judge nor to have the officers appear before the judge in order to accept their appointment. Likewise, the principles whereby experts may be rejected are debatable as well (see Art. 467 and ff, 662, 663 and 723 LECrim).

(ii) At the evidence proposal stage, to specify the name and surname of the expert who is going to make a statement on the case is no longer needed (Art. 656 LECrim). However, it is believed that it is necessary that the party concerned should propose the relevant evidence so that the judge may take into account in

his/her judgement the test results. Ex officio evidence is not allowed when article 726 LECrim (related to documentary evidence) establishes that "the Court itself will examine the books, documents, papers and other pieces of conviction…" This principle is contained in Art.728 LECrim (at a party request) and it makes possible ex officio evidence with the exceptions quoted in Art. 729 LECrim (Pardo, 2008).

(iii) The manner of giving evidence in Courts is modified as well: in the case of documentary evidence, the expert does not attend the trial to ratify his/her report or to answer/clarify any question related to such report. As noted above, this is a relevant point because it may not be easy to interpret the conclusions of the analysis. Moreover, it is possible that the report is not even read during the hearing (cf.: Art. 726 "the Court itself will examine…").

Additionally, the rejection of the analysis will not necessarily imply the presence of the experts during the hearing. In those cases where any party challenges the report SSC case law (see, e.g., SSC sentences dated September 27, 2006, July 2, 2008 and February 3, 2009) proffers several solutions depending on all possible circumstances. Thus:

(i) The experts will not appear during the trial if the party rejects the report without solid grounds. The analysis, however, will be assessed as evidence for the prosecution.

(ii) When the party rejects the report and proposes the expert's appearance without proper justification, the court will not accept it and once more the analysis could be assessed as evidence for the prosecution (Noncase law agreement ruled by the Spanish Supreme Court on May 25, 2005).

(iii) If the expert's appearance is requested based on solid grounds, the court will accept it if appropriate. Then, the Tribunal will assess the expert's statements.

(iv) Finally, it is possible that a party may reject the report and submit a contradictory report. Both reports should be assessed by the Court according to the rules of "sane criticism."

If, as a general rule, the evidence submitted which is sufficient to plead successfully the presumption of innocence is one undertaken according to the principles of orality, contradiction, immediacy and publicity, it seems obvious that converting the expert evidence into documentary evidence will imply the contravention of the publicity and immediacy principles, along with strong limitations on the contradiction principle as well.

Fifth, translating expert evidence into documentary evidence will have an impact on the appeals against judicial decisions. Thus, as regards the appeals to the Supreme Court (appeals for cassation) it needs to be considered that analyses on drugs are "documents" according to the second paragraph of Art. 849 LECrim. As regards appeals, it would also be important whether the analyses are considered documents or expert evidence (evidence of a personal nature). Note that judgements of acquittal and some aggravating convictions too will be almost impossible to observe when the mistake affects the assessment of evidence of a personal nature (Calderón-Cuadrado, 2005), for example in the Spanish Constitutional Court (henceforth, SCC) case law (see, e.g., sentence number 167/2002 on September 18, 2002 and another more recent sentence numbered 120/2009 on May 18, 2009).

In conclusion, by virtue of practical or operational factors, the legislature pursues a fiction: to change the nature of certain expert reports and turn them into documents, in spite of prejudice to judicial guarantees.

This reform is disagreed with for the aforementioned reasons. Additionally, the aim pursued, that is, to avoid experts attending the trials being able to thwart a negative effect on their work, could have been reached by other means while preserving the expert nature of evidence. In this regard, expert substitution is widely accepted, meaning that if the expert who carried out the analysis is not available, he/she can be replaced by another expert. Videoconference (65% of the appearances in court of the Civil Guard forensic experts to give evidence were carried out by videoconference in 2009) is another possible technique to allow the expert to make a statement without moving to the court facilities.

6.1.4 A recent drug analysis proficiency test by ENFSI

It has been observed that the same type of reasoning has been expanded to other criminalistic areas by Spanish Courts (see the next section), and it is possible that by case law or new law amendments, most or all expert evidence areas, when reports have their origin in official laboratories, could be reclassified as documentary evidence. Drug analysis was selected as a pioneering technique because of its “objectivity”, but even in this area controversy has arisen as a recent proficiency test from ENFSI on the nature and purity estimation of drugs has recently shown (details in the Appendix).

Experimental data variability is not only explained by the proper nature of what is being observed, but by the applied observation methodology among other variability sources. When the weight of drug samples is measured, the weight measurement uncertainty depends on errors due to calibration processes of the precision scale, to biases of the precision scale, and its own weight procedure. The ISO standard titled “Guide to the Expression of Uncertainty in Measurement – 1993” is recommended by the international scientific community to calculate the weight of drug samples by means of precision scales.

Consequently, any forensic expert wishing to weigh drug samples from seizures should also include uncertainties associated with their measurements. However, some drug official forensic reports have lacked uncertainty calculations, resulting in quite a lot of disparity in the results achieved among different European official laboratories, with some substances remaining undetected by some of them (for more information on the variability expected in the determination of the nature and purity of drugs from European official laboratories and from some other contributors involved in the ENFSI Drugs Working Group in a proficiency test held in late 2008 (ENFSI Drugs Working Group, 2008-2009), see the Appendix).

This example, anyway, was artificially scheduled by the organizers. Drug forensic cases have huge variability. Active ingredient quantification is the first stage for any analysis to be a success because, for example, a seized drug is often mixed with other substances or could be concealed; environmental or case conditions could bias the analysis results; uncertainty measurements from weight and purity analytical processes have to be correctly combined and interpreted; sampling criteria has to be clearly determined and justified, and so on. Consequently, the claimed objectivity in such analyses is not synonymous with easiness but rather; is full of intellectual obstacles. It seems to make no sense to deal with any drug case applying the same criteria, as those criteria should be explained or defended in court.

On the other hand, this example is shown to illustrate probable interpretation problems with the so-called factual opinions (AFSP, 2009, pp. 161-164). A drug forensic report should include information about homogeneities of large drug seizures, the type of range that could be expected, and whether the results obtained are much more probable if one alternative hypothesis is true (for instance, the probability of observing these quantitative results if the purity is above a certain level) over another one, following the likelihood paradigm. Therefore, it is not primarily the analytical result which it is important, but what it may mean in the context of the case.

6.1.5 Implications for other forensic examinations and possible future consequences

Since 2002, some case law have been made by different Spanish Courts applying the analogy of drug examinations, that is, the consideration as documentary evidence of psychological, psychiatric, and veterinarian forensic reports, fingerprint examinations, and forensic reports made by medical examiners (Pardo, 2008). Based on simple arguments which served to classify these other types of forensic examinations as “objectives”, and guaranteeing the independence of official forensic experts because they belong to public Institutions, scientific discussion on how to evaluate data as evidence was set aside. Practical reasons to expedite proceedings (a benefit for the laboratories as well since forensic experts do not need to travel to court as often to make statements) justify the support given to this change of judicial treatment of forensic reports by jurists, directors or senior officials of public laboratories, and even by forensic experts.

This policy may lead to a future application of this principle to any kind of expert report. Therefore, new legal reforms in that direction could be proposed. So, although the US scientific community tries “to promote the development of forensic science into a mature field of multidisciplinary research and practice, founded on the systematic collection and analysis of relevant data,” as Recommendation 1 of the National Research Council report (National Research Council, 2009) states, one of the fields in need of a greater improvement in daily forensic practice all over the world, that is, evaluation of evidence from a solid logical framework, is simply ignored. On the contrary, conclusions of official forensic reports are shielded and problems of interpretation of evidence are subordinated to the ability of lawyers.

6.1.6 Contrasting the US Supreme Court case of Melendez v. Diaz with Spanish law

In June 2009, the United States Supreme Court, in a case involving a certificate of analysis indicating the presence of an illegal drug, established as unconstitutional under the Sixth Amendment to the Constitution the admission in evidence of a certificate of analysis without the expert's testimony [Melendez-Diaz v. Massachusetts, 129 S.Ct. 2527 (2009)]. The question raised was whether a conviction may be sustained based upon an expert's certificate of analysis without the ability of the defendant to confront (cross-examine) the expert. The certificate of analysis was considered "testimony" by the Supreme Court, and therefore the defendant was entitled to question the expert about the report. If the defendant had the opportunity to question the expert under oath prior to trial, such as during a deposition, the presence of the expert at trial may not be required, and the certificate of analysis, along with a transcript of the deposition, could be admissible. In either event, the ability to question the expert witness would help insure that the expert's opinion on the nature of the drug would be neutral and objective.

The Spanish situation is completely different. It is not possible to come to the same conclusion. Art. 24 of the Spanish Constitution establishes some guarantees called constitutional rights including, as is pertinent to the case under study, actual legal guardianship, to be judged in due process of law, to use applicable means of proof for self-defense, and prohibition of defenselessness. However, the SCC, which is the interpreter of the Spanish Constitution, has warned that:

(i) The right to use means of proof is a right bestowed on the Spanish legislature to establish the regulating standards for its exercising (SSC case law, see, e.g., sentence number 136/2007, June 13, 2007). As a consequence, the legislature regulates the different evidence and may establish a series of limits and, in this regard, it may establish that reports on drug analyses may be regarded as documentary evidence.

(ii) This right is not absolute and unconditional (SCC case law, see, e.g., sentence number 11/1981, April 8, 1981). According to Art. 24 of the Spanish Constitution, only Judges and Courts can "examine the lawfulness and relevance of proof" (SCC case law, see, e.g., sentence number 60/2007, March 16, 2007). Therefore, though a party may request the presence of a scientific expert before the Court to give evidence, in case of being considered unnecessary by that Court, that presence will not take place and this fact will not be unconstitutional.

(iii) Not all types of irregularities in a criminal process may be regarded as a constitutional contravention, but only those which cause actual defenselessness (SCC case law, see, e.g., sentence number 15/2005, January 31, 2005).

(iv) Drug analyses have frequently been carried out before trial and constitute evidence existing before trial. If it is being formally submitted at the investigative stage of the criminal proceeding on an undisputed basis the judge himself may examine them (SCC case law, see, e.g., sentence number 15/2005, May 9, 2005).

As regards that said in the SCC case law, it is believed that the Spanish Constitutional Court – which should decide what is unconstitutional – considers that Art. 788.2 of LECrim regarding the nature, amount and purity of drugs conforms to the Constitution, and the failure to declare of experts in court cannot be regarded as breaching the constitution even if this expert report is the only evidence for the prosecution. However, that article refers to particular types of analyses (the drugs analyses carried out by official laboratories regarding the nature, amount and purity of said drugs) and limited to a specific process (the so-called brief criminal proceeding), not spreading its scope beyond strictu sensu.

6.1.7 A paradigm shift?

6.1.7.1 The origin of the problem in Spain (and most European countries)

The paragraph titled "Challenges Facing the Forensic Science Community" (National Research Council, 2009, p. S-3) reads as follows: "For decades, the forensic science disciplines have produced valuable evidence that has contributed to the successful prosecution and conviction of criminals as well as to the exoneration of innocent people… Those advances, however, also have revealed that, in some cases, substantive information and testimony based on faulty forensic science analyses may have contributed to wrongful convictions of innocent people. This fact has demonstrated the potential danger of giving undue weight to evidence and

testimony derived from imperfect testing and analysis. Moreover, imprecise or exaggerated expert testimony has sometimes contributed to the admission of erroneous or misleading evidence."

In spite of the advances of forensic science for decades, admission of erroneous or misleading evidence has taken place with possible wrongful convictions of innocent people. The instruction by the Congress of the United States to the National Research Council to undertake the study (National Research Council, 2009) (at the request of a Consortium of US Forensic Science Organizations) because it required significant improvements. Behind this need for significant improvements was also the previous controversies expressed by the academic world and the public opinion in the United States for years (Kennedy, 2003, p. 1625).

Therefore, one of the problems entailing that Europe is not equivalent to the US as regards the justification of the need for a report like that of the National Research Council report (National Research Council, 2009) comes from the absence of such a previous controversy in most European countries. It is worth mentioning, the initiative carried out in England and Wales by the House of Common's Science and Technology Committee during 2004 and 2005, publishing the Seventh Report titled Forensic Science on Trial. This reports states that the Association of Chief Police Officers had agreed on the need for a protocol regarding the validation of scientific techniques prior to their being admitted in court. The absence of controversy partially explains the very different reactions in the European laboratories since the release of this report, and the very different speeds to insert the new paradigm in daily forensic practice. The ENFSI Board scheduled a comprehensive program of activities to announce it during 2009, specifically at the 21st Annual Meeting held in Ankara (Turkey) in May 2009 and at the Seminar held on August 24, 2009 and August 25, 2009 at the Bundeskriminalamt in Wiesbaden (Germany). Nonetheless, the US report should be comprehensively studied within each European laboratory and in conjunction with each national judiciary. Perhaps results could be expected thereafter affecting the judicial scene.

To avoid the serious risk of minimizing the strength of the current sensitivity about the necessity of improving the evaluation of evidence in court, the academic and scientific communities should be involved in each country to guarantee a sound and effective discussion. The risk is highest in countries where almost all forensic reports are made by official institutions in criminal proceedings. It is for this reason that the academic and scientific involvement and/or the internationalization of the debate are absolutely necessary.

6.1.7.2 How to change the paradigm

On one hand, the controversy among statisticians along the last century has been quoted and its impact on science. On the other hand, mention has also been made of the relevance of many common ways of reporting conclusions and their consequences in Courts and even in statutory laws. "How can mentalities now be changed, tried and tested under old paradigms, mainly inside law enforcement agencies or prosecutor's offices long used to considering some scientific proof as infallible?" And "How can it be explained to the courts that until now some evidence was deemed safe, but from now on, the same evidence will be deemed uncertain?" In fact, one does not need to be a prophet to predict the kinds of reactions to be expected. The current situation in many laboratories is that it is not difficult to find generalized inconsistencies in the ways of reporting conclusions among different fields of expertise in the same laboratory, or within the same criminalistic field.

Recently, the Swedish National Laboratory of Forensic Science (SKL) announced in the aforementioned Vth Conference of the European Academy of Forensic Science its decision to unify the manner of drawing conclusions, regardless of the fields of expertise involved, in order to provide the same kind of information to courts (Nordgaard, 2009). It is believed that this should be the first step that every laboratory should undertake in order to change mindsets. And as previously done by this laboratory, the entire staff of forensic experts should be absolutely convinced of the need to do just that. The Swedish laboratory needed years to achieve this objective.

Other institutions have made important efforts to ensure that judges and prosecutors understand the new paradigm, to interpret evidence in trial (during recent years, a number of European Universities and Laboratories have started seminars, courses, conferences, and the like for judges and prosecutors on the evaluation of scientific evidence). However, there is no official information as to how the laboratories related with the corresponding law enforcement agencies have viewed this paradigm shift.

In the meantime, the opinion on how forensic laboratories should act as legal reforms on evaluation of evidence in criminal proceedings are being studied by jurists consists of reporting the need for a paradigm shift, which has been steadfastly supported by the scientific community for the last two decades. The National Research Council report (National Research Council, 2009) was a solid and strong call for a change in the traditional way of thinking in evidentiary data evaluation within the whole forensic community.

6.1.7.3 Possible alternative solutions for the Spanish Code of Criminal Procedure reform

The Spanish legal reform objectives have to be compatible with the demands of science. The statutory law, which was amended by adding a second paragraph to Art. 788, was preceded neither by any previous scientific debate nor any reference to the international state of the art. Spain runs the risk of distancing its evidence legislation from Europe if scientists do not take part in law commissions before reforming their criminal proceedings and if the debate is not immersed in the global forensic paradigm shift.

Following examples of other countries as (National Research Council, 2009; The Law Commission Consultation Paper No 190, 2011; AFSP, 2009, pp. 161-164), nations should be able to find good ways of adapting evidence regulation to current scientific state-of-the-art.

In the meantime, provisional solutions could be implemented such as incorporating technology and procedures which speed up proceedings. Of course, those measures alone are entirely insufficient. Legal reforms will be necessary, which preserve both judicial and scientific requirements.

6.1.8 Conclusions

In this paper, it has been endeavored to show why any legal reform dealing with scientific evidence in Court needs to preserve the legal and constitutional guarantees while align itself with state-of-the-art scientific knowledge. The very nature of forensic evidence reliability has been challenged in the last two decades, and a new paradigm is spreading steadfastly across countries towards a more solid and well-grounded forensic science. However, the 2002 Spanish Code of Criminal Procedure reform veered towards the opposite direction: objectivity and reliability were attributed to expert evidence provided by official laboratories by way of

the legal status of documentary evidence, instead of that of classical expert evidence and to official reports dealing with weight, purity and nature estimations on seized drugs, a status which could be expanded by case law or other legal reforms to other criminalistic areas. It is believed that the objectives of reform regarding the efficiency and simplicity of laboratory procedures for reporting and testifying in Court, can and must be met satisfying simultaneously both legal guarantees and international scientific standards.

Furthermore, any supplementary guidelines to the ISO 17.025 standard on accrediting identity criminalistic tests, or any policy to improve this type of forensic service should be based on a solid logical framework, appropriate to the efficient and scientific evaluation of evidence, as is the case in a DNA comparison. If not, those measures could bring discredit on the accreditation system in forensic science since transposed conditional fallacies could not only be committed but accredited!

APPENDIX I: ENFSI Proficiency Test on drug analysis.

In this test, there were four different sources for drug samples: A. Heroin mixture; B: Cannabis (Hashish); C1 and C2: 3-CPP tablets with traces of amphetamine; D: amphetamine mixture. The organizer asked participants for the nature and purity of drugs. Sample C had two different drugs, each of them in very different proportional quantity.

Homogeneous pieces of samples firstly by a "Grindomix" mill and secondly by mortar and pistol were distributed to all participants.

ENFSI had 54 members (almost all of them are European governmental laboratories) when the collaborative test was carried out. The organizer numbered up to 71 possible participants in the test (it is usually allowed to some nonmember of ENFSI laboratories to take part in such as activities).

The main techniques used by participants were high-performance liquid chromatography with diode-array detection (HPLC-DAD) to quantify active ingredients, and gas chromatography-mass spectrometry (GC-MS), in the main, to detect the nature of them.

The organizer summarized results providing the following data per sample: mean, standard deviation (SD), relative percentage between standard deviation and mean (RSD), median, highest and lowest value when numeric values were given by participants (purity); and quality information ("+": hit; "-": fail; "(+)": hit with some objections) when were not given those numeric data (nature). Table A.1 summarizes purity results from the four tests (some errors in SD and RSD of Sample C1, and RSD of sample C2 were detected by the authors of this paper, and were corrected, see Table A.2). Box plots are shown in Figure A.1.

TABLE A.1 - Descriptive statistics for the variables

Purity	Sample A	Sample B	Sample C1	Sample C2	Sample D
No. of labs	56	48	30	15	55
Mean	32,5	5,33	14,85	0,93	13,99
Median	33,4	4,95	14,85	0,4	14,2
SD	3,98	1,64	2,16	1,56	1,77
Highest	39,8	12	21,8	6,3	18,2
Lowest	9,6	2,6	11,6	0,1	7,9
RSD	12,2%	30,7%	14,5%	167,8%	12,7%

TABLE A.2 - Descriptive statistics for the variables once outliers removed

Purity	Sample A	Sample B	Sample C1	Sample C2	Sample D
No. of labs	51	44	28	13	52
Mean	33,06	5,03	14,42	0,44	14,12
Median	33,4	4,88	14,70	0,4	14,25
SD	1,93	0,81	1,43	0,22	1,34
Highest	37,03	7	16,73	0,8	16,99
Lowest	27,93	3,7	11,6	0,05	10,9
RSD	5,8%	16,1%	9,9%	50%	9,4%

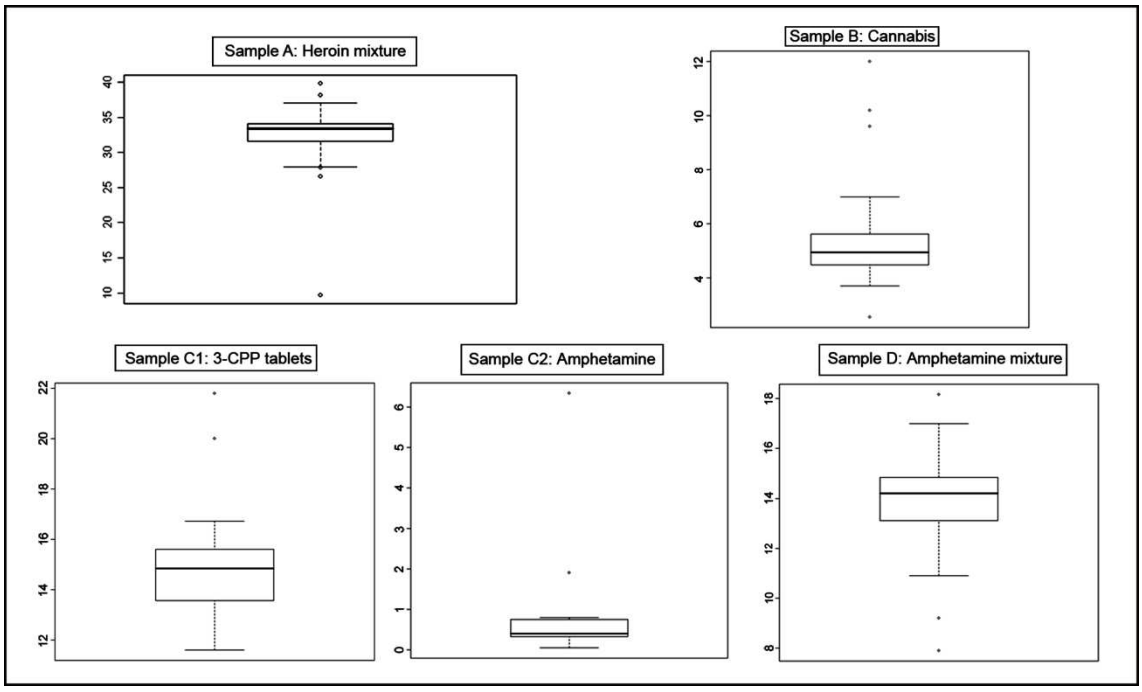


FIGURE A.1-Box plots showing median (thicker black line) and outliers

Chapter 7. TECHNICAL SUPPORT FOR A JUDGE WHEN ASSESSING A PRIORI ODDS[8]

Abstract

The Engineering Department of the Spanish Civil Guard has been using automatic speaker recognition systems for forensic purposes providing likelihood ratios since 2004. They are quantitatively much more modest than in the DNA field. In this context, it is essential a suitable calculation of the prior odds in order to figure out the posterior odds once the comparison result is expressed as likelihood ratio. These odds are under the responsibility of a Judge, and many consider unlikely that they can be quantitatively calculated in real cases. However, our experience defending in Court over 500 speaker recognition expert reports allows us to suggest how the expert may support Judges from a technical point of view to assess the odds.

Technical support as referred should be preferentially provided in the preliminary investigation stage, after the expert report being issued by the laboratory, as in the course of oral hearings it is much more difficult for those who are not familiar with the new paradigm. It can be initiated upon request by the Examining Judge or any of the litigant parties. We consider this practice favourable to the equality of arms principle.

The use of Bayesian networks is proposed to provide inferential assistance to the Judge when assessing the prior odds. An example of the explanation above is provided by the case of the terrorist attack against Madrid-Barajas Airport Terminal 4 perpetrated in December 2006.

Keywords: BN in judicial contexts, prior odds, evaluation of evidence, likelihood ratio, speaker recognition.

7.1 Introduction

Besides the DNA expert reports there are hardly any other fields of criminalistics in which the Spanish official experts use Bayesian inference to scientifically interpret the results achieved by their analyses. And in the DNA expert report, the use of this inference in practice does not seem –in our opinion– to be an example to heed. If this was the case –and bearing in mind the years which have elapsed in which DNA techniques have been deployed in Justice- the majority of Spanish Judges and Courts would have to have a clear idea of the difference between the logical basis of the conclusions of the DNA report –in which assertions never appear about the authorship of the traces as can be found in the reports of handwriting or fingerprints– and those from other fields of criminalistics in which it is necessary to carry out a comparison between known and unknown source samples. As it has been stressed before, our personal perception in this regard is that we are still far from deploying the Bayesian inference correctly (Lucena et al., 2012, pp. 952-963).

An outstanding exception in Spain is that of the Engineering Department of the "Guardia Civil" police force (from now on Spanish Civil Guard) as regards to voice comparisons and audio or video recording authentication. The Department's Acoustics' Area has specific software for evaluating the likelihood ratio in a voice comparison developed by the AGNITIO company, the upshot of long uninterrupted, constant scientific research since 1997 with the university research group Area de Tratamiento de Voz y Señales (ATVS) belonging to the Further

[8] LUCENA MOLINA JJ, GASCON ABELLAN M, PARDO IRANZO V, "Technical support for a judge when assessing a priori odds". Law, Probability & Risk, 14(2), June 2015, pp. 147-168.

Polytechnical School of the Madrid Autonomous University and since 2004 with said company. As regards authentication, the Department assigns subjective probabilities using expert knowledge to inform likelihood ratios (O'Hagan, 2006). Undoubtedly, this less known way of assessing the scientific evidence could be the immediate solution to avoid making the fallacy of the transposed conditional in conclusions (Aitken, 2004, pp. 79 ff.).

Within the logical framework of Bayesian inference, the prior odds are required –as well as obtaining likelihood ratio– to ascertain the posterior odds, in other words, the information really resolves the needs of the Courts as regards to voice comparisons: calculating the posterior probabilities of the propositions put forward by the parties in the process regarding voice authorship questioned in the light of the results of the comparison and the context information at the disposal of the court (Aitken, 2004, pp. 79 ff.). The expert's aid is not only used to calculate likelihood ratios but it also might be used to quantify the prior odds and, on this latter assumption, it is specifically the court which has the information required so that the expert can do it.

However, during a BBfor2[9] (Bayesian Biometrics for Forensics) Workshop held in Martigny (France) in December 2011, N. Brümmer said in a presentation the following: "Let's be realistic: In a real court case, it is unlikely that numbers will be assigned to the prior odds and cost ratio, and that the posterior will be explicitly calculated." It's believed that most forensic experts would agree with such a comment, but is there no room for possible solutions?

The purpose of the paper here is to discuss what some scholars have proposed in the past as guidance (Section 7.2.1), the role of priors in a forensic evidence comparison within a Continental legal system as the Spanish one (Section 7.2.2), a case study illustrating why and how to use Bayesian networks to calculate priors (Sections 7.2.3 and 7.3.1), a more deeper study as regards to expert assistances to courts in terms of assessment of scientific evidence (Section 7.3.2), and conclusions (Section 7.4).

7.2 Material and methods

7.2.1 The Analysis of Evidence School

The work entitled 'Analysis of Evidence' by Terence Anderson, David Schum and William Twining (Anderson, 2010), is regarded as a classic in the rational assessment of evidence within the Anglo-US legal world, and could be a good starting point to face the challenges. These authors developed the thinking of John Henry Wigmore (1863-1943) who stated that neither common sense, intuition, nor experience, all of which are parts of the exercising of the skills inherent in inferential reasoning can replace what he called 'principles of proof'.

These experts in the rational assessment of evidence have known how to describe a sound rational structure which helps to understand how, based on the evidence (whether tangible or physical or testimonial) a proposition may be proven which they call the 'ultimate probandum' (the last thing which needs to be proven) whose content is related with legal descriptions of a certain crime. A recent manuscript written by Paul Roberts and Colin Aitken (Roberts, 2012) on the logic of forensic proof says the following: "Wigmorean method is nothing more (or less) than an attempt to summarise the logic of inferential reasoning in graphical form, tailored to specific intellectual (analytic and decision) tasks. It is, in other words, a practical heuristic for litigation support designed specifically to assist those who need to formulate, evaluate, or

[9] http://bbfor2.net/ (accessed on January 12, 2017).

respond to arguments to inferring factual conclusions from mixed masses of evidence to improve the quality of their intellectual output."

Using the graphics' method (Figure no.1), this reasoning method can be organised into a scheme whose key constructor elements are the concepts of proposition, generalization and evidence.

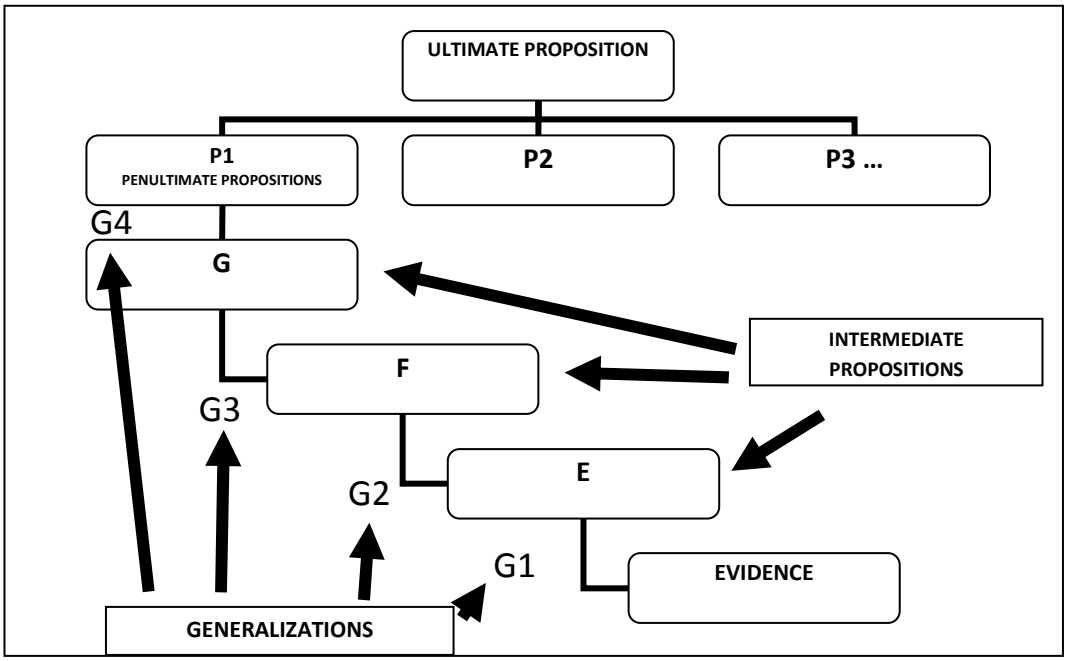


Figure no. 1 – Scheme used in Wigmorean methodology

This methodology will be used in this article to make it easier for readers on how to understand the important distinction between 'evidence' and 'background information' in the case study (see Section 7.2.3), that is, certain propositions will be considered as 'evidence', and others as 'background information'[10]. The background information will be introduced in the notation using the capital letter I (see Section 7.3.1).

It is worth pointing out that although Anglo-Saxon and Continental legal systems have notable differences, what is set out in this paper goes beyond them, as it is related with the reasoning logic bases which must be the background of these systems to be classified as rational.

7.2.2 The role of priors in forensic evidence comparisons

In order to make a decision at the end of the criminal proceedings it is first of all necessary 'to prove facts', those which are relevant to the criminal prosecution, and afterwards to apply to those facts the legal consequence according to the law for such a kind of facts. Let us remember that though it is said 'to prove facts' for our own convenience, in fact what are

[10] This is an important point because on one hand "when a proposition's degree of belief is evaluated, there is always exploitation of available *background information*, even though it is not explicit" and, on the other hand "A *relevant proposition* is taken to mean a proposition which is not included in the background information", and besides it should be assumed that, for the evaluation of forensic evidence, "the distinction between the background knowledge of the Court or of the expert, and the findings submitted for judgement, is clear from the context'' (Taroni, 2006, pp. 3-4).

proved are not facts, but statements on facts (Gascon, 2010, p. 76). The proof of statements on facts is based on information provided to the proceedings through evidence. We call 'prior probability' the subjective probability assigned to a statement on facts bearing in mind certain information or piece of evidence at that moment.

Priors of any statements on facts are changing as the proceedings are going on. Therefore, there will be different priors depending on the moment of the procedural stage. Some of them could be useful to justify investigative actions (e.g. phone tapping, house search, e-mail examination, and so on). Once has the trial stage finished, the court has to determine which facts have been proven and justify their factual decisions in the sentence.

There are two stages in the Spanish Code of Criminal Procedure: the investigative and trial stages. The former is mainly addressed to finding out the circumstances around a crime and its authorship. During this phase, investigative actions (e.g., scene examination or questioning) and precautionary measures (e.g., pre-trial custody), are undertaken, but these activities do not actually constitute legal evidence and therefore they cannot invalidate the presumption of innocence, nor be enough to convict. At the second stage, under a different Judge, the prosecutor makes a decision to charge and the evidence is provided.

Experts can help Judges in both stages according to the Spanish Code of Criminal Procedure. In accordance with case law, a Judge cannot yield or relinquish his / her evaluating responsibility. Judges are responsible for determining prior and posterior odds of the Bayes' Theorem in a real case, although they can be assisted by experts to do that scientifically (maintaining 'coherence' in probabilistic reasoning (Lindley, 2007)) and combine both of them with likelihood ratios. While the investigative stage allows experts to explain much better the procedure and conclusion of their reports to both the investigating Judge and parties, the trial is the exclusive legal stage in which the report will be converted into scientific evidence. Though the Court is made up of different magistrates from those who acted in the first stage, the parties are the same. Therefore, the assistance given by experts quantifying prior odds in the investigative stage will reinforce the principle of equality of arms in the proceedings (Champod, 2010). Besides, the expert will never give such assistance to the Court without knowing the alleged relevant information submitted by the parties in the proceedings, something that only Judges can determine and authorize. A wider and specific legal study as regards to expert assistance in terms of the appraisal of a scientific test will be presented in Section 7.3.

7.2.3 Case study: a voice comparison

As a practical example of the application of the graphic method proposed by Wigmore, we have based ourselves on the Report 08/10279/AI-01 by the Acoustics and Image Department of the Spanish Civil Guard (voice comparison) and information taken from Sentence 18/2010 of Section 3 of the Criminal Division of the High Court dated May 21, 2010 regarding the investigation into the attack which occurred at parking D of Terminal T-4 of the Madrid-Barajas Airport on December 30, 2006[11]. The graphic (Figure no.2) may help the Court to carry out an assessment of prior odds in favour of the prosecution thesis before the forensic voice comparison provides its technical information. At the root of the graphic, art and science have to be combined as its creators specifically recognise. The quoted example is not intended to be anything other than an academic example –illustrative – because in reality it should be completed by the evidence, propositions and generalizations deriving from the defence.

[11] http://en.wikipedia.org/wiki/2006_Madrid-Barajas_Airport_bombing (accessed on January 12, 2017).

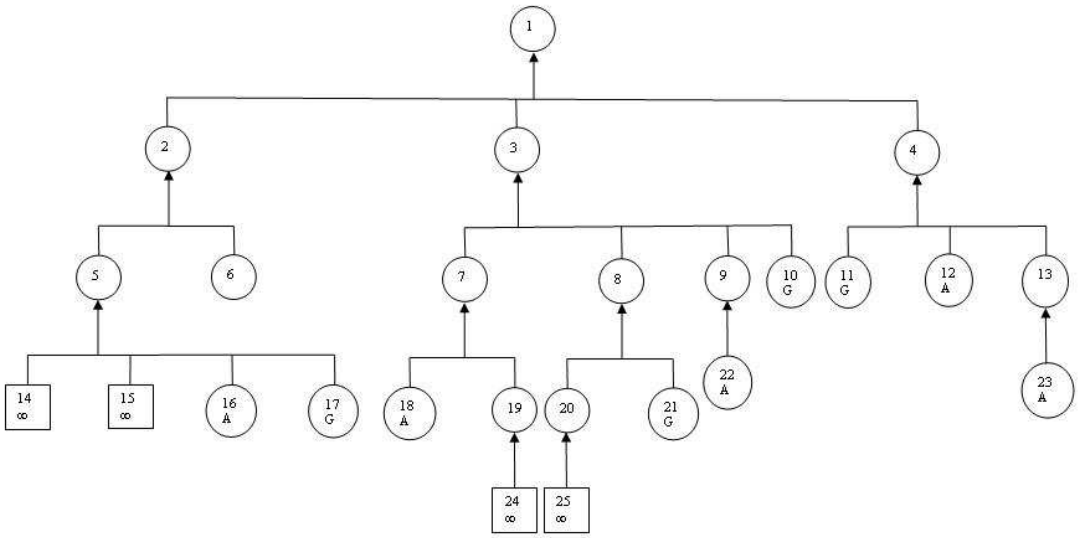


Figure no. 2 – Graphic of the case study

KEY:

1: IP made a phone call to the DYA Exchange in S. Sebastián, in ETA's name, at 7:53 a.m. of December 30, 2006 from the mobile phone XXX warning that a van bomb had been planted with a large amount of explosives at parking D of T-4, being a Renault Traffic model, dark red, registration plate YYY which would go off at 9:00 a.m.
2: IP made a phone call to the DYA Exchange in S. Sebastián at 7:53 a.m. of December 30th 2006 from the mobile phone XXX.
3: IP was aware that the Renault Trafic van bomb, dark red colour, YYY registration plate, was loaded with big quantities of explosives, which was parked at parking D of T-4, and would explode at 9:00 a.m.
4: IP was aware that it was acting in the name of ETA.
5: IP bought the phone XXX.
6: The S. Sebastián DYA received a call from the phone XXX at 07:53 a.m. of December 30, 2006.
7: IP knew that the van bomb was loaded with a large amount of explosive.
8: IP knew that the van bomb was parked at parking D of T-4.
9: The explosion was scheduled for 9:00 a.m. on December 30, 2006.
10: The warnings regarding the planting of bombs by ETA in public places are usually carried out with a sufficient time margin for an urgent evacuation and preventing deactivation by putting at serious risk the deactivators' safety.
11: ETA usually claims the more notorious attacks in similar media shortly after the occurrence of the facts.
12: ETA claimed the T-4 attack in a press release published in the daily paper GARA in its digital edition dated January 9, 2007 and in the normal edition dated January 10, 2007.

13: The upshot of this handwriting comparison supporting the authorship by IP of the handwriting of the telephone number of the Madrid Fire Brigade in a note seized from IP clothes when he was arrested.
14: Statement by MS saying that IP bought the phone with the number XXX.
15: Statement by the Phone vendor describing the physionomical characteristics of the purchaser of the phone XXX on December 23, 2006.
16: Invoice for the purchase of the phone XXX dispatched by Phone vendor on December 23, 2006.
17: The miscreants use prepaid phones so as not to leave any trace of their involvement in the committing of any offences over the phone.
18: The Renault Trafic van, dark red in colour, registration YYY was stolen in France by hooded individuals on December 27, 2006. His driver was released by 9:40 a.m. of December 30, 2006.
19: The Renault Trafic van, dark red in colour, registration plate YYY, was collected by MS and MSS, with the explosive charge already prepared for activation on December 29, 2006.
20: IP took part in the operation thought up for the transfer of the van bomb to T-4.
21: The members of a terrorist command co-ordinate with each other to undertake an attack.
22: The explosion of the van bomb took place at 08:59:29 a.m. on December 30, 2006.
23: IP and MS were arrested by the Spanish Civil Guard in Mondragón (Guipúzcoa) in a road check on January 6, 2007. In the IP's napsack two guns stolen in France on October 26, 2006 were found. IP's personal documentation contained a handwritten note with the phone number of the Madrid Fire Brigade and the S. Blas Police Park (Madrid). The Madrid Fire Brigade received a call from the phone XXX at 7:55 a.m. of December 30, 2006 warning that the van bomb had been placed at T-4.
24: Statement by MS describing the collection of the van bomb in France on December 29, 2006.
25: Statement by MS describing the method of operation for moving the van bomb to Madrid-Barajas Airport.

LEGEND:

DYA: Road Assistance Service of the Basque Country.
ETA: Euzkadi Ta Askatasuna.
IP: terrorist no. 1.
MS: terrorist no. 2.
MSS: terrorist no. 3.
T-4: Terminal 4 of the Madrid-Barajas Airport.

SIGNS:

∞: testimony which the Court can hear directly in the courtroom or evidence which may be directly inspected in the same way.
A: proposition assumed to be right.
G: generalization.
€: witness evidence.
O: circumstantial evidence or inferred propositions.

The graphics' method helps to reason with logic and it proves particularly useful and even necessary when the case is complex. However, it is not suitable to calculate prior odds in accordance with the laws of probability and Jeffrey's rule, laws and the rule which ensure rationality when estimating such odds without any doubt.

In addition to the information provided in the previous graphic by Wigmore –limited so as not to make the example too lengthy – it is worth stressing that the Court is aware that:

- The members of the ETA command kept an appointment with their managers MGAR and JAA who proposed to them to take a van to T-4, requesting that they should first check the route. In actual fact, to undertake the work route survey they rented a vehicle in Irún (Guipúzcoa) in the name of IP (the Court has the rental contract for the vehicle Volkswagen Polo registration VVV, signed in Irún in the name of IP for October 21, 2006 and October 22, 2006). The vehicle travelled a distance of 963 Km (the distance between Irún and Barajas is 928 km).
- The van bomb was stolen in France at gunpoint and its owner – Spanish – was released by an ETA command 40 minutes after committing the terrorist attack on the T-4.
- The ETA command police record from its inception given by the Information Service of the Spanish Civil Guard.

All this information (far from what was described known by the Court in the case in hand) can be used to evaluate – in odds form – the propositions defended by the parties to the process with regard to the voice comparison: that it was IP who made the phone calls on the mobile XXX on December 30, 2006 warning of the planting of the van bomb in contrast to that it was not done.

We can use a Bayesian network to evaluate the prior odds so that in this way – as has already been indicated previously – the laws and theorems of the probability theory and Jeffrey's rule are respected. Furthermore, we are interested in exploring to what extent these odds can be changed when the Court considers some of the information on which the evaluation of the odds is based right.

To this end, the first task which the expert has to carry out in this regard is to consider the propositions which must be inserted in the network and the relations between them and we will do so showing the following network as shown in Figure no. 3 (it is a very simplified network to promote the teaching purpose of the article):

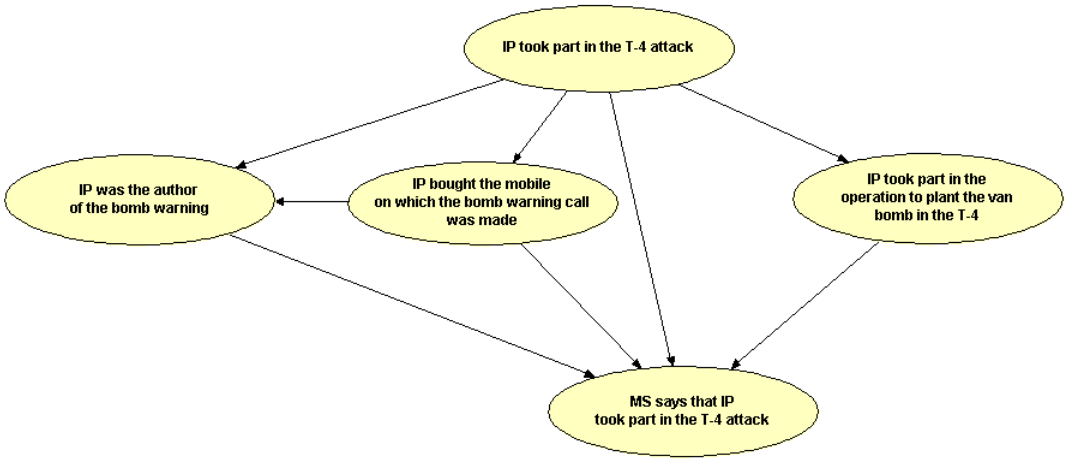


Figure no. 3 – Network of the case study

The propositions are defined inside the ellipses which we call nodes. From top to bottom we discover what we could call the 'ultimate proposition': the participation of IP in the terrorist act perpetrated at T-4 of Madrid-Barajas Airport. The possible statuses of this proposition would be as follows: principal / accessory / non-participation – including in this status merely material or innocent participation if possible. This proposition is said to be 'relevant' for the four directly linked to it in descending order:

a) that IP carried out the bomb warning to the DYA exchange (questioned voice inspected by the Spanish Civil Guard);
b) that the mobile use in the bomb warnings was bought by IP;
c) that he actively participated in the transport of a van bomb from Irún to T-4; and
d) that the statement by MS about the participation of IP in the T-4 attack was true.

'Relevance' is one of the main concepts used to construct Bayesian networks (Taroni et al., 2006, p. 3). It is said that a proposition B is relevant for another A if and only if the answer to the following question is positive: if it is supposed that B is true, does that supposition change the degree of believe in the truth of A? The relevance, as can be gleaned from the definition provided, is commutative: if A is relevant to B, B is to A. This commutative property of relevance is the necessary consequence of the axiom which constitutes the third law of probability.

Henceforth capital letters will be used instead of describing the enunciations of the propositions on the network with a view to facilitating the use of mathematical notation when developing argumentation.

The new network thus has the following appearance:

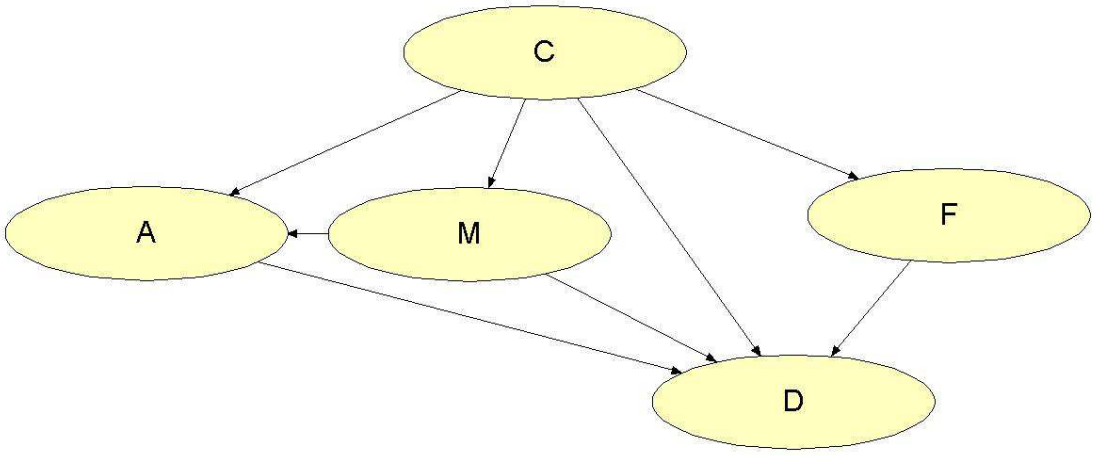


Figure no. 4 – Network of the case study to facilitate the use of mathematical notation

We can illustrate the commutatively of relevance by concentrating at the nodes C and M, for example. Knowing that IP took part in the T-4 attack is relevant to establish whether it was he who bought the mobile used to make the bomb warning and, on the other hand, knowing that he bought the mobile used to make the bomb warning is relevant for knowing whether he took part in the T-4 attack.

On a network only the direct dependencies between the nodes should be reflected. On the network in Figure no. 4 not all the nodes are linked to each other. Nodes C and D are linked to all of them but, for example, F is only linked to C and D and M to C, A and D. This network has the virtue of containing all the possible ways of making connections between nodes: 'divergences' (between node C and nodes A, M, D and F; and between node M and nodes A and D); 'series' (between nodes C, M and D; between C, A and D; between C, M, A and D; and between C, F and D); and 'convergences' (between nodes A, M, C, F and node D; between nodes C , M and node A).

The Bayesian networks use the property of conditional independence to optimise the calculation of the probabilities which may prove to be of interest within the network. The connections of Figure no. 4 allow the assurance of conditional independence where necessary.

The network of Figure no. 4 has a high degree of interdependence: if IP had taken part in the T-4 attack, the network propositions in intermediate position could be related with his criminal activity. On the other hand, the statement by MS, very lengthy on details, contains the intermediate propositions. This leads to the fact that the network connects the node C to all the others and the node D as well. While the former connects by means of a divergence, the latter is achieved by means of a convergence.

The divergences between node C and nodes A, M, D and F complies with that which is believed that if you do not know the status of C, the knowledge of the state of one of the other nodes (A, M, D or F) may change the belief about the probabilities of the statuses of the other nodes which are in brackets. For example, if it is not known for sure whether IP took part in the T-4 attack, knowing that he acted as the driver of a shuttle vehicle when moving the van bomb to parking D of T-4 from Irún may change the belief about whether it was him who provided the bomb warning, bought the mobile used in the warning or whether the statement by MS was right. We could provide the same reasoning knowing A with regard to M, D and F, knowing M with regard to A, D and F or knowing D with regard to A, M and F.

Analysing the compliance of the property of the conditional independence in the divergence between C and nodes A, M, D, F the following occurs (whenever the status of node D is not known) if the status of C (C1, C2 or C3) is known:

- a change in the belief of the statuses of F does not cause any change in the belief about the statuses of nodes A and/or M and vice versa.

Convergence indicates that if there is a certain idea whether any of the uses of the statement by MS (D) is known definitely, a change in the belief of any of the statuses of nodes A, M, C and F brings about a change in the statuses of the others.

With the aid of the Hugin programme[12], it is possible to verify all the comments about the conditional independence and dependence in the network of Figure no. 4.

It is very important to establish that on Bayesian networks only uncertain propositions can be included (in the form of nodes). Everything about which there is confidence it has occurred cannot be introduced therein. It is a logical reasoning structure under uncertainty.

[12] http://www.hugin.com (accessed on January 12, 2017).

7.3 Results and legal discussion

7.3.1 Running the Bayesian network

The next step in our argument consists of checking the usefulness of the network for calculating prior odds about whether IP made the bomb warning. To illustrate that it is not necessary to deploy figures to prior odds, we commence our probabilistic reasoning by comparing probabilities with each other, simply considering to what extent some may be bigger than others.

Exclusively based on general information about the case (we designate this information with the letter I), in other words, the information which the Court may have about whether IP may have committed the T-4 attack with the absence of any of that written on the network, we can determine the following statuses of the proposition C: 'IP took part in the T-4 attack', mutually exclusive and exhaustive[13]:

- C1: as principal in the first degree
- C2: as accessory or principal in the second degree
- C3: did not take part or did so innocently

It is possibly not hard to be in agreement with the following assignment of prior probabilities from the background information explicitly mentioned in the Section 7.2.3: $P(C1 \mid I) > P(C2 \mid I)$. Let us imagine that the Court initially is just in the sceptical doubt: $P(C1 \cup C2 \mid I) = P(C3 \mid I)$, in other words: $P(C1 \mid I) + P(C2 \mid I) = P(C3 \mid I)$. The Court has no reasons to believe more the prosecutor hypothesis than the defence one. These probabilities are those of the propositions in view of the information of a general context.

We can make numerical assignments to the different statuses of proposition C, respecting the aforementioned qualitative assignment of probability (Table no. 1):

C:	C1	C2	C3
P(C | I)	0.45	0.05	0.5

Table no. 1

The numerical values actually chosen serve as guidelines and are justified in the light of the general information available about the case by the Court. The most important is not the numerical assignment chosen for each status, but rather the numerical proportions between these assignments. As these are mutually exclusive and exhaustive events, the sum of their probabilities must be equal to the whole number to meet the two first probability theory laws. These figures indicate that the Court estimates there is a 50% probability that IP took part in

[13] Art. 28 of the Spanish Penal Code (SPC): Principals are those who perpetrate the act themselves, alone, jointly, or by means of another used to aid and abet. The following shall also be deemed principals: a) Whoever directly induces another or others to commit a crime; b) Whoever co-operates in the commission thereof by an act without which a crime could not have been committed.
Art. 29 SPC: Accessories are those who, not being included in the preceding Article, co-operate in carrying out the offence with prior or simultaneous acts.

the T-4 attack and that, in turn, it is believed to be more likely that he did so as an offender rather than as an accomplice.

Other probabilities which we can estimate in qualitative form are those technically called likelihoods. These are probabilities of events considered on the network, in other words, of the propositions which are under the ultimate proposition which we have called C since it is regarded as right any specific state of this proposition and, where applicable, other propositions. As an example: P(A | C1), P(A | C2) … P(M | C1) … P(F | C1) … P(D | C1) … P(D | A, M, C1) … P(D | A, F, C1) … P(D | A, M, F, C1)…

The likelihood that IP took part in the attack (C), having bought the mobile used in the bomb warning (M) and in view of the information of a general context known by the Court (I), can be expressed as follows as regards to each of the statuses of proposition C:

$$P(M \mid C1, I) = P(M \mid C2, I) > P(M \mid C3, I)$$

In this regard, we believe that the likelihoods of the C1 (offender) and C2 (accomplice) statuses are the same as regards to the fact that IP bought the mobile used (M) and the general context information (I). The context information (I) causes the likelihood of the participation of IP in the attack to be estimated as greater than that of an innocent action if we consider it correct that IP bought the mobile (M). Furthermore, we must bear in mind that the ETA command is made up of three members of the band (we regard this information as integrated in I) any of them may have carried out this duty. Finally, it is also borne in mind that the mobile purchase may have been carried out by someone collaborating with the ETA command or even some innocent person who has misplaced the mobile or had it stolen.

A numerical assignment respecting the previous considerations may be shown as given in Table no.2:

C:	C1	C2	C3
M: P(M = t | C, I)	0.2	0.2	0.1
M: P(M = f | C, I)	0.8	0.8	0.9

Table no. 2

In the same way, the likelihood of whether IP took part in the attack (C), having given the bomb warning to DYA (A), purchased the mobile from which the warning was made (M) and the general context information (I) may be expressed in the following way for each of the possible C statuses:

$$P(A \mid C1, M, I) \geq P(A \mid C2, M, I) > P(A \mid C3, M, I)$$

In this case, we have borne in mind the fact that the mobile which made the bomb warning was available, after buying it, constitutes an additional opportunity as regards to being the warning offender. Furthermore, we know the following from the general context information I – which the purchase of the mobile can preferably be assigned to IP by means of a deduction based on a physiognomic description of the vendor of the mobile used to carry out the bomb warning and the known physical features of the other members of the ETA command. Based on the assumption that it is considered that IP did not buy the mobile, the criteria set out in the assignment of the first of the likelihoods are considered, though downward numbers are

written owing to the fact that it is regarded as less probable that there was a collaboration consisting of carrying out a bomb warning than purchasing a mobile – bearing in mind the information available about how the events occurred.

If we translate the above in numerical assignment, Table no. 3 respects the aforementioned qualitative assignments:

M:	t			f		
C:	C1	C2	C3	C1	C2	C3
A: P(A = t | C, M, I)	0.8	0.4	0.01	0.1	0.1	0.001
A: P(A = f | C, M, I)	0.2	0.6	0.99	0.9	0.9	0.999

Table no. 3

As regards to the likelihood that IP took part in the attack (C), acting in the van bomb planting process (F), and in view of the general information about the case (I), the following would seem reasonable:

$$P(F \mid C1, I) > P(F \mid C2, I) > P(F \mid C3, I)$$

It has been estimated that common sense and the experience of previous attacks by the terrorist organisation ETA would incline us to think that the complexity of the van movement operation to its placement at parking D of T-4 would require something more than mere occasional cooperation on the committing of the attack. Even so, this possibility is in no way ruled out and complicity is considered. However, it is considered that $P(F \mid C3, I) = 0$, as it is no longer regarded as reasonable to argue that IP would have taken part in the movement operation of the van bomb unless he was an accomplice, at least, to commit the attack.

Translating the previous considerations into numerical form, we get the following results as shown in Table no. 4:

C:	C1	C2	C3
F: P(F = t | C, I)	0.8	0.2	0
F: P(F = f | C, I)	0.2	0.8	1

Table no. 4

The latest likelihoods which are of interest to us to estimate on the network are as follows:

$$P(D \mid C1, M, A, F, I) > P(D \mid C2, M, A, F, I) > P(D \mid C3, M, A, F, I)$$

The qualitative assignments carried out are justified bearing in mind that the likelihood that IP took part in the attack as the offender (C1) gains strength compared with the other propositions (complicity or non-participation) if the explanatory facts M, A and F are regarded as correct. In the same way, there is a reflection on the advantage of complicity proposition (C2) compared with non-participation (C3) if such events are regarded as correct.

The numerical assignments we can carry out become complicated in this case because the number of variables to be borne in mind and their statuses mean that the number of

assignments required attains 24 (with their corresponding assignments of complementary probabilities).

The probabilities which it is wished to calculate regarding the truthfulness of the statement by MS, once the values of the statuses of the other variables and their complementarities are known. The criteria in selected order are as shown in Table no. 5 and the 48 assignments are shown in Table no. 6:

P(D = t | M, F, A, C)	Variable statuses
0.999	When all the binary restrictive events have the value 'true' and the variable C = C1
0.99	When all the binary restrictive events have the value 'true' except for M and the variable C = C1
0.9	When all the binary restrictive events have the value 'true' except for A and the variable C = C1; also on the assumption that all the binary events considered have the value 'true' and the variable C = C2
0.8	When C = C2 and A and F have the value 'true'
0.7	Both when C = C1 and only F has the value 'true' as well as when C = C2 and only A has the value 'false'
0.6	when C = C2 and only F has the value 'true'
0.3	When C = C1 and only F has the value 'false'
0.2	Both when C = C1 and only A has the value 'true' as well as when C = C2 and only F has the value 'false'
0.1	When C = C1 and only M has the value 'true' and when C = C2 and only A or M have the value 'true'
0.01	When C = C3 and only F has the value 'false'
0.001	When C = C3 and only A or M have the value 'true'
0	When all the other restrictive events have the value 'false' regardless of the status of C and when C = C3 and F has the value 'true'

Table no. 5

C:	C1								C2								C3							
A:	t				f				t				f				t				f			
F:	t		f		t		f		t		f		t		f		t		f		t		f	
M:	t	f	t	f	t	f	t	f	t	f	t	f	t	f	t	f	t	f	t	f	t	f	t	f
D: P(D = t | M, F, A, C)	0.999	0.99	0.3	0.2	0.9	0.7	0.1	0	0.9	0.8	0.2	0.1	0.7	0.6	0.1	0	0	0	0.01	0.001	0	0	0.001	0
D: P(D = f | M, F, A, C)	0.001	0.01	0.7	0.8	0.1	0.3	0.9	1	0.1	0.2	0.8	0.9	0.3	0.4	0.9	1	1	1	0.99	0.999	1	1	0.999	1

NOTE.- In the numerical assignment of probabilities, the general criterion has been established of not exceeding an assessment which entails more than three decimals.

Table no. 6

The usefulness of the network is appreciated when in the Hugin programme, the Edit mode (the mode in which the nodes and their links are devised and the initial probabilistic values are completed in the tables) passes to the Run mode (the information flow is activated in the network). The example network has been shown below when the Run mode is commenced (the probabilistic values are those of each proposition):

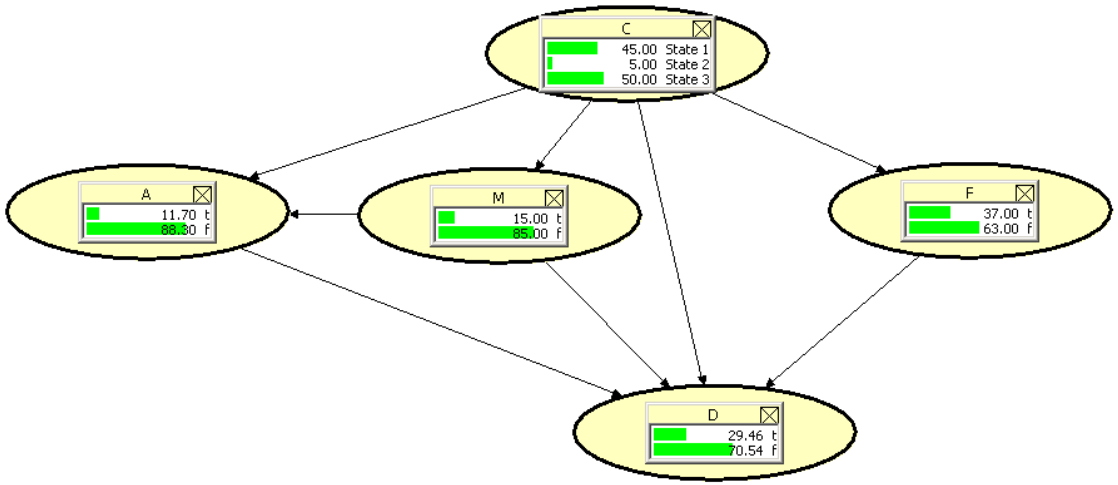


Figure no. 5 – Starting the run mode

We can carry out an initial Bayesian inferential reasoning exercise with the information we have available:

- With the general context information available (I), the probabilistic estimates carried out, at the first instance, of the explanatory facts which appear on the network and in accordance with the probability laws, an initial estimate can be made that there is a 11.7% estimate that IP could be who gave the bomb warning.

However, the network provides us with the possibility of applying Jeffrey's rule (Taroni, 2006, p. 20), in other words these are related to the changes in the information flow within the network. For example, if the Court was certain that IP took part in the operation to plant the van bomb, it could be estimated how much influence that belief would have on the probability which we wish to ascertain: that IP gave the bomb warning. Let us observe the results as shown in Figure no. 6:

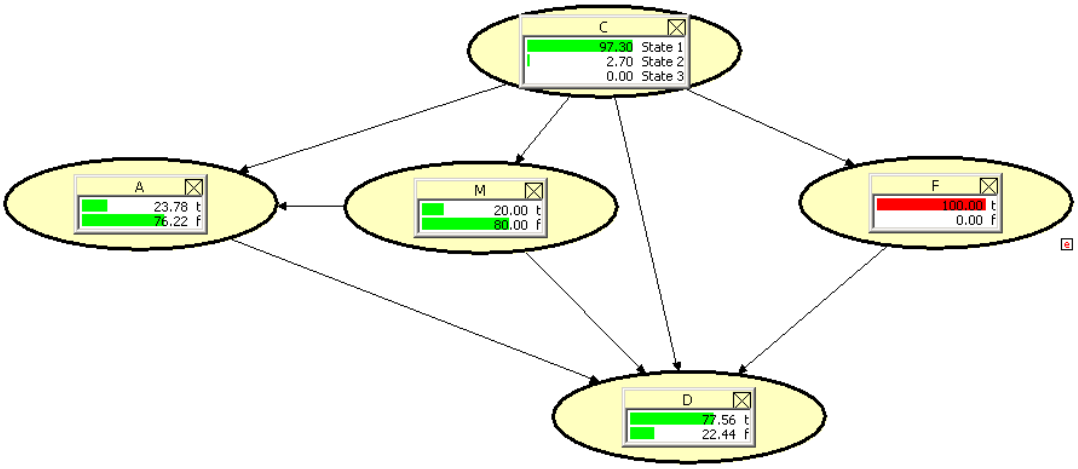


Figure no. 6 - Instantiating node F

The influence of node F has not been highly significant in terms of increasing the probability that IP was who gave the bomb warning. Even so, the network estimates that this probability increases up to 23.78%. It is worth observing that the probability that IP took part in the T-4 attack increased from around 45% to about 97.3%.

Maybe we could assume that it could be more important for our purposes that the Court should be certain that IP bought the mobile on which the bomb warning call was made. This being the case, the results are as follows in Figure no. 7:

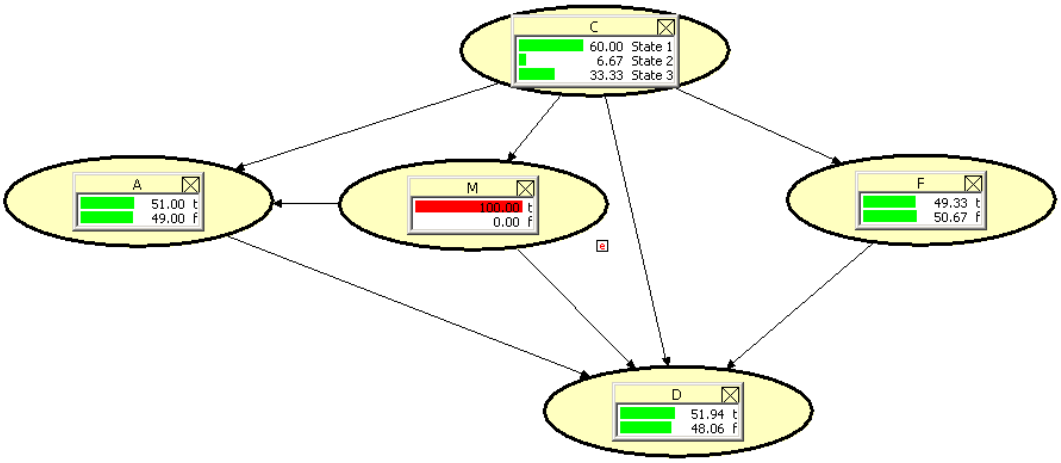


Figure no. 7 – Instantiating node M

Against this backdrop, the increase in the probability we are seeking is much more important than before: we attain 51%. However, the probability that IP took part in the attack rose from 45% to 60%.

And if we consider that the Court estimated that the events M and F are right, in what way would these certainties influence which we have in hand? (Figure no. 8):

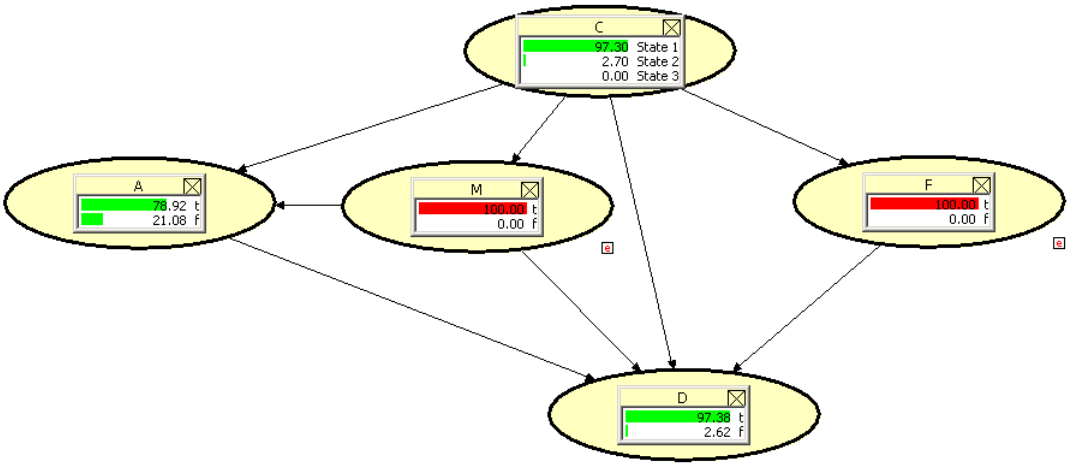


Figure no. 8 – Instantiating nodes F and M

The rise has now been much more important: we almost reach 78.92%. The probability that IP has participated in the attack has attained 97.3%, the same as the figure obtained with just bearing in mind the certainty about F. What has risen a lot is the probability that the statement by MS is right: it has reached 97.38%, starting at 29.46% in the initial Run mode status.

Finally, on the assumption that the Court was right about what was stated by MS, we would obtain the following result:

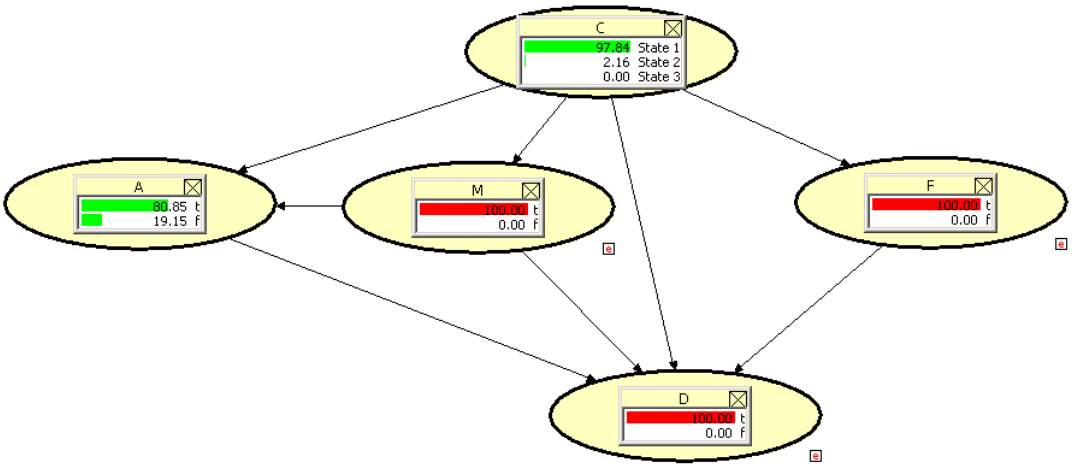


Figure no. 9 – Instantiating nodes D, F, and M

In this way a figure of 80.85% has been attained. In turn, the probability of C1 has been estimated at 97.84%.

Carrying out a summary exercise, the Court could argue in favour of prior probability that IP had carried out the bomb warning that around 80% of the assumption which considered the propositions looked at in the network right except for the so-called 'ultimate proposition' in this network (node C). However, the Court could adopt a more realistic position – which in this case would favour the accused party – considering that the explanatory events featuring in the network could never be entirely demonstrated and that the statement by MS – like no other evidence of a testimonial nature – may not be exact. In Hugin language it is called entering likelihood evidence and in this case Hugin performs updating on uncertain evidence.

The Bayesian network only endeavoured to help the Court with the calculation of the prior odds before the expert in voices had intervened. The results of their analysis will modify the aforementioned initial bet and allow the probabilities to be found which are being searched for. In our example we tried to find the probability of IP were the author of the bomb warning once the result of the voice recognition was obtained.

The experts of Spanish Civil Guard in voice comparison who acted in court regarding the attack on T-4 argued, before the Court, that to properly interpret the information provided by them in the expert opinion about the voice comparison ('Evidence shows a moderately strong support towards the hypothesis which points at the defendant IP as the source of questioned voice versus the hypothesis of the authorship by other people recorded in similar acoustic conditions', wherein the likelihood ratio was equal to 213 and was used the Evett verbal scale) required a calculation of the prior odds within the Bayesian inference framework used by them to attain an estimate about the probability of attributing to IP the bomb warning. Making it clear to the Court that this prior odds was an exclusive competence of the jurisdictional body, they wished to illustrate this as carrying this out in practice if the members of such a body had the confidence that one of the command was the one who made the call. The experts wished to stay out of the actual case in point and there were 4 members of the so-called 'legal command' (as has been seen, in reality the ETA command was made up of only 3 'legal' members – non-fugitive members of the terrorist band from Spain). In this way, IP was able to attain 25% of a priori probability of being the author thereof. As can be seen, this figure can always be classified as conservative with regard to the maximum calculated with the network.

Applying the odds form of Bayes' Theorem, the calculated posterior probability that IP was the author of the bomb warning once the result of the voice expertise had been attained, having approximately reached 98%. This figure is what stood out in the media as provided by the experts after their appearance in court to defend their expert report, but the patient reader of this article will have noticed that the journalists were far from reality as they had properly understood the information conveyed by the Spanish Civil Guard experts.

If we work out the posterior odds in favour of the proposition that IP made the bomb warning based on the maximum a priori probability for this proposition with the Bayesian network, the figure of around 99.9% would be reached for the posterior probability that IP was the author of the bomb warning.

This was not set out in the provisions of the sentence regarding the voice test be regarded as brilliant. Without any doubt, there is a difficulty of understanding both in that stated by the experts in trial as well as that written in the expert report.

7.3.2 On the expert assistance to Courts in assessing forensic expert testimony

The assessment of evidence is at the core of the judicial decision-making. It consists of determining whether the evidence before the Judge or Court supports (and how strongly) the statements about disputed issues of fact, which implies to evaluate the reliability and helpfulness of the available evidence as well as to ascribe appropriate weight to it in support of such statements. Assessing evidence consists, in other words, in determining, through a chain of 'reasoning' from the evidence submitted in the courtroom, whether the statements about disputed factual issues can be taken as true.

So far as evidential reasoning is essentially made up of inductive inferences based on probabilistic laws or even on social generalizations without any scientific foundation, the results that said reasoning produces cannot be taken as conclusive but rather they must be read in terms of simple 'probability' (Gascon, 2010, p. 104). Self-evidently the above does not mean that such results cannot be treated as true; what's more, there are good reasons to expect that the results of a sound process of assessment will be accurate. This only means that, strictly speaking, the most that can be said is that the evidential reasoning concludes with a 'hypothesis', a statement which we accept as true though we do not know with absolute certainty whether it is or not, and that the degree of probability will provide a good criterion for its justification. The consequence of the just mentioned is clear: the main task of a rational assessment is that of 'measuring probability'. And that's why the objective of the evidence assessment models must be to provide rational schemes to assign the probability of the hypotheses.

Two major models for assessing evidence may be adopted (Gascon, 2010, pp. 92 ff.): a) those based on confirmation schemes, which use the logical or inductive notion of probability (which applies to propositions and corresponds to the common use of 'probably', 'possibly', 'presumably' something is true); and b) those based on the application of mathematical tools to the appraisal process, which use the mathematical concept of probability (which applies to events and it is interpreted either in terms of the relative frequency of the class of events to which they belong or in terms of the degree of belief in the occurrence of an event or in the veracity of a proposition). The first model is by far the one prevailing in the legal field but there

are also abundant approaches which propose to apply the mathematical methods of the probability calculus to the evidence assessment process.

This work is actually inserted into the second model: the use of the Bayes' Theorem as a tool to rationalise the assessment of scientific evidence; in other words, to 'rationally determine what is to be believed' about the hypothesis to be proved when a scientific test has been carried out (Royall, 1997; Gascon et al., 2010). A task that –we should remember– is undertaken by combining the result of the expert report on the occurrence of the disputed hypothesis compared with its alternative (which is expressed by means of a 'likelihood ratio') with the 'prior odds' of these hypotheses in view of all the evidence that the judge has before him apart from (and on the sidelines of) the expert report.

However a remark must be made in this regard. Bayesianism, as a 'general' model for evaluating the support that all the available evidence provides to a hypothesis on disputed issues of fact has been subject to criticism[14]. Most of the criticism is driven by practical considerations. In particular they highlight how difficult it may be for the judge to quantify its prior appraisal (the 'prior odds'); or draw attention to the inevitable subjectivity of said quantification, or to the risk which entails putting it in the hands of judges statistical instruments which most of the times are incomprehensible to them: perhaps not all the experts – it is argued – would actually be capable of presenting the 'likelihood ratio' with the necessary accuracy for its correct understanding; moreover, even if they were able to, the judges could still get confused[15]. Furthermore, the Bayesian formula – they say – entails a relatively simple mathematical calculation in the base case of having to evaluate a 'single' item of evidence which 'directly' deals with the hypothesis to be tested, which in turn is a 'simple' hypothesis. However the difficulty of the calculation increases strikingly when it is used to resolve more complex situations such as the 'plurality' of evidence relating to a hypothesis, the 'cascaded inference' (or indirect evidence) or the proof of a hypothesis relating to a 'complex' fact.
To counter this criticism the following should be noted. First, the criticism stated does not take into account the advances which have been made in the last few decades to optimise and simplify the probabilistic calculations (for example by means of the deployment of Bayesian networks) and it actually reveals the traditional aversion of jurists to open up to extralegal knowledge. Secondly, this paper is not focused on the adoption of the Bayes' Theorem as a general evidence assessment model, but rather only as a way of determining (by quantifying) the probative value to be assigned to the statistical data supplied in the forensic expert report. We consider that to be particularly appropriate as in recent years the evolution of 'forensic science' has provided a very wide range of scientific proofs with a statistical structure which can be used in all the proceedings[16]. Finally, what is specifically proposed here is the use of a tool (the Bayesian networks) to quantify Court's prior subjective assessment (the 'prior odds'),

[14] For the deficiencies assigned to Bayesianism applied in criminal procedural law, see (Gascon et al., 2010a). See also (Tribe, 1971; Taruffo, 2004; and Cohen, 1977).
[15] The main problem here is the possible confusion of P (E/H) with P (H/E) and P (E/no-H) with P (no-H/E). It is really hard for experts to express the proposition «P (E/H) is much bigger than P (E/no-H)» in a verbal form which does not make it sound like «P (H/E) is much bigger than P (no-H/E)». That is why those who defend the use of this appraisal model in the courts consider that it will not attain its purposes whilst judges and lawyers are not familiar with the most elementary probability theory. See (Friedman, 1996, pp.1836-37, and footnote number 9).
[16] In actual fact, the *Bayes' Theorem* constitutes the logical reasoning structure to resolve the appraisal problem that the specialized scientific forensic community has claimed as the most appropriate for decades. See (Fienberg, 1991).

which is fairly – as mentioned above – one of the main difficulties indicated by critics in order to use the Bayes' Theorem in Courts. And there can be no objections to this proposal: there is no reason to assume that the result of this operation is of higher quality when the judge carries it out 'subjectively', without the aid for reasoning provided by the logic and the mathematics.

There is no doubt that the use of Bayesian networks is –as it has been endeavoured to show here– a good way to rationally quantify the 'prior odds'. But there is no doubt either that, in practice, use of this tool requires the assistance of expert staff, which in principle may appear as 'problematic'. The objection could be raised that, by using such assistance, the risk is run that it is the expert and not the Judge who ultimately determines what is to be believed about the disputed factual issues under consideration, thus converting the expert into the decision-maker of the case. In that way, however, -here is the 'problem'– the principle of 'free' assessment of evidence, whereby it is for judges and no one else to assess the evidence, would be breached and a system based on the Judge's authority would have been replaced by another based on expert's authority.

In our opinion though, there would be no grounds for the stated objection. The assistance provided by the expert to the judge is 'merely technical' and consists of using his knowledge to help the latter, turning his degree of belief into figures; in other words, quantifying that which he is unable to quantify by himself. That is all. It is the Judge who has to provide the pieces of evidence that form the basis for this degree of belief which has to be quantified. More specifically, contrary to the conclusions stated in the above-mentioned expert report (see Section 7.2.3), which are the entire responsibility of the expert as it is something which only he can reach with his tools and knowledge, the 'prior odds' (which –we should recall- expresses the degree of belief assigned by the Judge to the 'ultimate probandum' in odds form before the expert test is carried out and the results thereof are known) is something which is entirely based on the knowledge that only the Judge has; and, in actual fact, the Judge could find it himself if he just had the necessary training to do that. The expert merely helps him to quantify this degree of belief in probabilistic fashion.

Precisely because the expert aid provided to the Judge in this task is merely technical, this cooperation does not entail that it is the expert who ends up carrying out a function (that of judging and previously stating the facts of the case) which has been institutionally granted to the Judge, nor is it at odds with the principle of free assessment of evidence; quite the opposite, in fact. This principle has been traditionally interpreted in extremely subjective and irrational terms. In actual fact, it has been customary –and unfortunately it still is– to understand 'free assessment' as a private, firm, non-transferable, subjective conviction; as a kind of ineffable 'quid', a hunch which cannot be exteriorized nor controlled. The widespread expression 'intimate conviction' clearly expresses this irrational way of conceiving it. However this interpretation proves to be unacceptable. The Court's assessment of evidence must indeed be free from legal obstacles but clearly 'not free from reasons'. Otherwise we would be proclaiming the purest judicial decisionism. This is why when a scientific test has been carried out, the aid of the expert to calculate the 'prior odds' and determine from that juncture –using the Bayes' Theorem– the probative value which should be assigned thereunto, not only does it not prove incompatible with the principle of free assessment of evidence but rather it is even 'required'. There would be no point in deploying the complex expertise that technical-scientific advances put at the disposal of the proceeding if the determination of its probative value were totally left in the hands of an unskilled Judge who does not know its correct interpretation and scope. The expert assistance provided to the Judge who has to assess the evidence is thus

highly important to rationally determine the value of the 'prior odds'. Moreover, it could be said to constitute an 'epistemological guarantee' required by any rational conception of the assessment of evidence. It's something else deciding 'when' and 'how' this assistance can be provided.

The Spanish criminal procedural model, which is adversarial, is structured around two major stages, both of them assigned to different judges or magistrates[17] because the Constitutional Court Judgement dated on July 12, 1988 declared it to be unconstitutional for the same court to hear the case and pass sentence on it –to ensure the court`s impartiality-.

The main aim of the first stage – preliminary investigation -, as far as possible and never at any price, is to search for the truth. Acts are carried out which are intended to establish something which is not known (the circumstances relating to the offence and the identity of their perpetrator – article 299 LECrim (Spanish Code of Criminal Procedure)); its implementation would result, amongst other aspects, in the Judge's decision as regards to the alternative dismissal-opening of the trial. These actions thus do not, as a general rule, have any evidential value nor do they serve as the basis for any sentence regarding the guilt or innocence of the accused.

During preliminary investigation the Judge enjoys broad powers to agree, at the request of either party or on his own initiative, such diligences as he sees fit, including expert reports which he will agree 'when, to establish or assess any important circumstance or fact in the summary, scientific or artistic knowledge proves necessary or appropriate' (article 456 LECrim). Once the expert has provided expertise, the Judge may – also on his own initiative or at the request of either party – ask the experts such questions as he deems relevant and/or ask them for the necessary clarifications (article 483 LECrim).

Two things become clear that being said up to now:

a) Leaving aside the debate about the nature of the expert report and assuming that the expertise is just another means of proof or diligences of the investigation –depending on the procedural stage we are at – what is for sure is that the expert helps the Judge in his function, providing the process with knowledge that the magistrate does not possess. In this regard, inter alia, the TS (Supreme Court) Judgement on May 5, 2010 states that: expert reports contain 'technical opinions about certain matters or about certain facts by those who have special training in this area with a view to making the Court's job easier when appraising the evidence. It is not evidence which provides factual aspects but rather criteria which help the magistrate to interpret and appraise the facts without changing its attendant in order to appraise the evidence'.

b) The Judge's powers at this stage to request an expert report and to ask the expert those questions and clarifications the Judge deems opportune are broad. Should this be the case, self-evidently there is no legal impediment -at the request of the Judge and bearing in mind, obviously, that it is the latter (or the requesting party) who states 'clearly and decisively to the experts the object of their report' (Article 475 LECrim)– to the expert also helping him with the work to establish the 'prior odds' required within a logical framework of Bayesian inference to

[17] It is important to carry out this precision because in many countries around us the preliminary investigation falls within the competence of the Public Prosecutor. In Spain the only exception is assumed in the Criminal Procedure for Minors (LO 5 issued on January 12, 2000).

obtain the 'posterior odds'. Let's bear in mind that the expert will issue his report with regard to facts –to be precise, statements about factual data which he is informed.

The second stage, the trial, is the one at which, inter alia, the evidence is presented with a view to delivering judgement. This is subjected to the principles of oral procedure, contradiction, immediacy and publicity, but not by chance, but rather because they constitute the mechanism to more appropriately ensure the rights of the defendant, the powers of the accusers and the correctness of the court decision. The elements required to justify the sentence must be acquired by the court in the trial.

The above, applied to the expertise, means that for the expert report carried out during the preliminary investigation [18] to have evidential value, there needs to be a statement by the experts at the trial[19]. And here, although the material powers of the Judge are much more limited, there does not seem to be any impediment for the Court to request the expert to provide clarifications about the answers given to the questions posed by the parties or even to formulate new questions. Nevertheless, we deem it crucial to bear the following in mind:

a) Firstly, there is no doubt that as regards to the expert's testimony, there is no article like Article 708 LECrim –regarding the testimony– which specifically enables the president of the Court to ask the witnesses those questions he sees fit to ascertain the facts regarding which statements are being made. By contrast, Article 724 LECrim regarding the expertise indicates that the experts will answer the questions and the cross-examination of the parties without any reference to the Judge. Nevertheless, and always considering questions conducive to ascertaining the facts about which statements are being made, it makes no sense for there to be a different treatment between both types of proof[20].

b) Secondly, we are aware that in criminal proceedings, the evidence presented in trial is that requested by the parties which the Judge deems to be relevant (article 728 LECrim); evidence requested on the initiative of the Court has an exceptional nature and is limited to the assumptions set out in article 729 LECrim. Number 2 of the previous provision enables the Judge to 'agree the examination of evidence not proposed by either party which the Court deems to be necessary to prove any of the facts which have been subject to qualification documents'. This section has been redefined by the TS (TS Judgements dated December 1, 1993), establishing that evidence requested on the initiative of the Court is solely possible if aimed at verifying whether the evidence about the facts are reliable or not from the perspective of article 741 LECrim (which is known as proof upon proof).

[18] Though, by way of exception, the carrying out of the expertise in a trial stage may be agreed, it is more common to be carried out during the preliminary investigation.

[19] In this regard two points must be pointed out: firstly, the law – Article 788.2 LECrim - has 'converted' into a document some specific expert reports (the analysis about the nature, quantity and purity of narcotic substances carried out by official laboratories), attributing evidential value to the document which includes the report without said statement by the expert at the trial. On the other hand, without legal backing other expert reports (in addition to those legally known as scientifically objective) are sometimes receiving this same treatment, at least when they are not challenged. In both cases there is sometimes some doubt as to whether the right to defense is being fully respected (see Pardo, 2008; Lucena et al., 2011).

[20] In actual fact, in trials with a jury, jurors are specifically allowed, through the presiding judge and subject to a relevance statement, to pose to the experts in writing such questions as they deem fit to ascertain and clarify the facts with which the evidence is connected (article 46 Organic Law 5 enacted on May 22, 1995 regarding Courts with a Jury).

It thus seems that those who have great powers can also make use of lesser ones. In other words, if, for the purposes of proving any of the facts which have been subject to qualification documents the Judge is able to agree the examination of evidence, he has all the more reason to ask questions or clarifications from the experts who, proposed by the parties, have spoken at the trial.

7.4 Conclusions

This research provides reasoning to justify the assistance of forensic experts to Judges and Courts to assign prior odds when the evaluation of scientific evidence has been made by likelihood ratios. We consider it very relevant that the judicial reasoning used to defend a result related with a criminalistic comparison is based on logic laws which are used in science and not in hunches.

The mathematical tool proposed for such assistance is a Bayesian network, because it is a logical structure of reasoning under uncertainty based mainly on the Bayes' Theorem and Jeffrey's rule. These rules ensure that the subjective estimation of the probabilities required in the inferential process strictly respects the probability laws. No one can reasonably say that the probabilistic estimations which we have come to are arbitrary, unfounded, unscientific, intentionally skewed or random.

The prior odds which have to be estimated in a Bayesian inference framework related with an expert report can be assigned numerically in line with generalizations and common sense in many cases. A real case has been discussed in detail to illustrate how to implement in practice the inferential assistance suggested.

Legal implications of the proposed expert assistance to Judges and Courts in a Continental judicial system as the Spanish one have also been discussed. In accordance with the applicable Spanish law and jurisprudence we believe that there are grounds to assist them in this way. On the one hand, this technical support could be thought as an 'epistemological guarantee' to preserve a rational conception of the assessment of scientific evidence. On the other hand, this kind of forensic practice could reinforce the principle of equality of arms in the legal procedure.

Future practical research will be needed to make good use of the possibilities that Bayesian networks can offer to Judges and Courts (and the parties to the proceedings) to make inferences, not only to assess coherently the evidence but also to make decisions in the same way (i.e., using utility functions) along the procedural stages. This work is only a first of such possibilities.

Chapter 8. PERFORMANCE OF LIKELIHOOD RATIOS CONSIDERING BOUNDS ON THE PROBABILITY OF OBSERVING MISLEADING EVIDENCE[21]

Abstract

In this article we introduce a new tool, namely 'Limit Tippett Plots', in order to assess the performance of likelihood ratios in evidence evaluation including theoretical bounds on the probability of observing misleading evidence. To do that, we first review previous work about such bounds. Then we derive 'Limit Tippett Plots', which complements Tippett plots with information about the limits on the probability of observing misleading evidence, which are taken as a reference. Thus, a much richer way to measure performance of likelihood ratios is given. Finally, we present an experimental example in forensic automatic speaker recognition following the protocols of the Acoustics Laboratory of Guardia Civil, where it can be seen that 'Limit Tippett Plots' help to detect problems in the calculation of likelihood ratios.

Keywords: forensic statistics, likelihood ratio, misleading evidence, Tippett plots, Limit Tippet plots.

8.1 Introduction

The statistical evaluation of the evidence by means of likelihood ratios (LR from now on) is increasingly proposed for the evaluative interpretation of results in forensic science (Aitken et al., 2011, pp. 1-2). In this context, measuring performance of LRs is critical in the process of validating statistical interpretation methods prior to its use in casework (Ramos et al., 2013a, pp. 156-159). One of the effects that are associated with bad performance of LRs is 'misleading evidence', defined as evidence which has a LR in favour of the wrong proposition, i.e. evidence which has a LR ratio higher than one when the defence proposition (H_d) is true or smaller than one when the prosecutor proposition (H_p) is true (Hacking, 1965; Royall, 1997; Royall, 2000, pp. 760-780; Aitken and Taroni, 2004; Ramos et al., 2013a, pp. 156-159; Ramos et al., 2013b).

As Aitken and Taroni stated "a change in the odds in favour of the prosecution's proposition, through the value for the evidence different from 1, is a change in the probability of the prosecution's proposition" (Aitken and Taroni, 2004, p. 197). Therefore, the Court could be misled by making wrong decisions if misleading evidence is provided by the forensic examiner.

Misleading evidence is one of the most degrading factors in the performance of LRs, and should be somehow measured in order to evaluate its importance (Ramos et al., 2013b). Moreover, the presence of strongly misleading evidence, namely LR values that support the wrong proposition with a value much greater or much smaller than 1, is even more important and degrading. Therefore, although the LR framework for evidence evaluation is logically correct and does not need to be validated, the implicit or explicit consideration of misleading evidence and strongly misleading evidence of the particular models used for LR computation is of capital importance in order to check the validity of LR procedures prior to its use in casework.

[21] LUCENA MOLINA JJ, RAMOS CASTRO D, GONZALEZ RODRIGUEZ J, "Performance of likelihood ratios considering bounds on the probability of observing misleading evidence". Law, Probability & Risk, 14(3), September 2015, pp. 175-192.

In this work we review and analyse the concept of misleading evidence from the statistics literature. In particular, we focus on theoretical work that derives bounds on the maximum proportion of misleading evidence that can be observed in a set of LRs, which follows the concept of 'the probability of observing misleading evidence' (Royall, 1997). Then, we apply these concepts to regular performance measures in LR-based evidence evaluation, such as Tippett plots. We propose the use of so-called 'Limit Tippett Plots', where the aforementioned bounds on the proportion of cases yielding misleading evidence in a set of LR values is explicitly represented. With 'Limit Tippett Plots', we add valuable information to regular Tippett plots: the violation of the theoretical bounds of misleading evidence is explicitly shown, revealing problems in likelihood ratio computation. This represents a major improvement over Tippett plots, where there is not a baseline performance to compare with. Therefore, 'Limit Tippett Plots' can be used to detect, e.g. inadequate statistical models, bad selections of populations, etc. Moreover, a freely available software in Matlab™ has been provided by the authors in order to easily draw 'Limit Tippett Plots', which can be downloaded at http://arantxa.ii.uam.es/~dramos/software.html.

In this article, an experimental example is presented in order to illustrate the usefulness of 'Limit Tippett Plots', following the methods used by the Acoustics Laboratory of Guardia Civil (AL-GC from now on) in daily casework. There, 'Limit Tippett Plots' are used in order to show that, if some populations are badly selected in order to compute the LR, whether for selecting wrong models, or feeding the models with inappropriate numerical values, the bounds of the probability of misleading evidence are violated. This could not be noticed by using Tippett plots, but it becomes easy to see with 'Limit Tippett Plots'. In the experimental example, the protocols and databases followed by Guardia Civil are used, showing that the proposed performance representation is useful in a real operational environment.

This work is organized as follows. First, we review the concept of misleading evidence according to statistical literature. Then, we present the results about theoretical bounds on the probability of misleading evidence contributed by (Royall, 2000, pp. 760-780). Then, we propose and describe 'Limit Tippett Plots'. The aforementioned experimental example in forensic automatic speaker recognition is then presented. Finally, some conclusions are drawn.

8.2 Misleading evidence

A definition of 'evidence' in statistics can be given according to the so-called 'law of likelihood', as follows:

> "If hypothesis A implies that the probability that a random variable X takes the value x is $p_A(x)$, while hypothesis B implies that the probability is $p_B(x)$, then the observation X = x is evidence supporting A over B if and only if $p_A(x) > p_B(x)$, and the likelihood ratio, $p_A(x) / p_B(x)$, measures the strength of that evidence" (Hacking, 1965).

In forensic science, the LR paradigm exactly follows the law of likelihood, and therefore LRs express the strength of some evidence according to a pair of mutually exclusive propositions.

The next step is to define the concept of probability of misleading evidence. Using common notation in LR-based evidence evaluation, we define $P_p(A) \equiv P(A|H_p)$ as the probability of A given H_p, the prosecution proposition. Conversely, we define $P_d(A) \equiv P(A|H_d)$ as the probability of A given H_d, the defence proposition. Given a real value k > 1, we will call P_d (LR >

k) as the probability of having LR values greater than k, when H_d true. Alternatively, P_p (LR < 1/k) is the probability of, when H_p true, having LR values smaller than 1/k. These probabilities consider the variation in the evidence, and therefore it does not refer to the probability of the LR in a given case with a fixed observation; but the probability that, due to variation in the value of the evidence, the LR will be out of some bounds.

In the above definition, the value of k determines the strength of the misleading evidence. For instance, in (Royall, 1997, p. 25) the value from which the strongly misleading evidence is considered is k = 32 although this is just a convention assumed by the author.

The LR with continuous data is defined as:

$$V = \frac{f(x \mid H_p)}{f(x \mid H_d)} \equiv \frac{f_p(x)}{f_d(x)} \tag{1}$$

where $f(x \mid H_p) \equiv f_p(x) \equiv f_p$, when x is a value of the continuous variable X, and $f(x \mid H_d) \equiv f_d(x) \equiv f_d$, in the equation have originated the observation x of the continuous variable X. Notice that in forensic evaluation of the evidence the value of x is usually fixed, and the LR is computed from that observation. In this research, we will consider the distribution of x with respect to the observation of misleading evidence, and therefore the probabilities P_d (LR > k) and P_p (LR < 1/k) consider the variation of the possible values of x. As we will see below, misleading evidence can be formally addressed by the use of LR methods to evaluate the evidence in forensic science.

In (Royall, 2000, pp. 760-780) several theoretical bounds for P_d (V > k) and P_p (V < 1/k) are derived. These bounds constitute the motivation of this research. Thus, detecting violations of such theoretical bounds will mean that some problems may have happened in likelihood ratio calculations. The sources of those problems may be, e.g. a bad selection of models, the use of inappropriate data, etc.

8.3 Theoretical bounds of the probability of strong misleading evidence

8.3.1 The universal bound

According to (Royall, 1997; Royall, 2000), and (Aitken and Taroni, 2004), for any given pair of distributions f_p (x) and f_d (x) in (1), there is a bound of 1/k on the probabilities of strong misleading evidence, i.e. $P_d(V > k) < 1/k$ and $P_p(V < 1/k) < 1/k$. Therefore, for a given likelihood ratio V computed in operational conditions, if $P_d(V > k) > 1/k$ or $P_p(V < 1/k) > 1/k$, then a problem has happened in LR calculation. More details about the universal bound and its justification can be found in (Royall, 2000, pp. 760-780). It is important to note that this universal bound applies to any statistical distribution, no matter its kind or shape, or whether it is parametric or non-parametric.

The universal bound 1/k is obviously a function of k, namely the value of strong misleading evidence whose probability is being computed. Figure 1 shows such an upper bound as a function of k. It is observed that the universal bound on the probability of misleading evidence will be decreasing with k. This is because if k' < k, the values of V representing strong

misleading evidence for k are also strong misleading evidence for k', and the opposite might not be true.

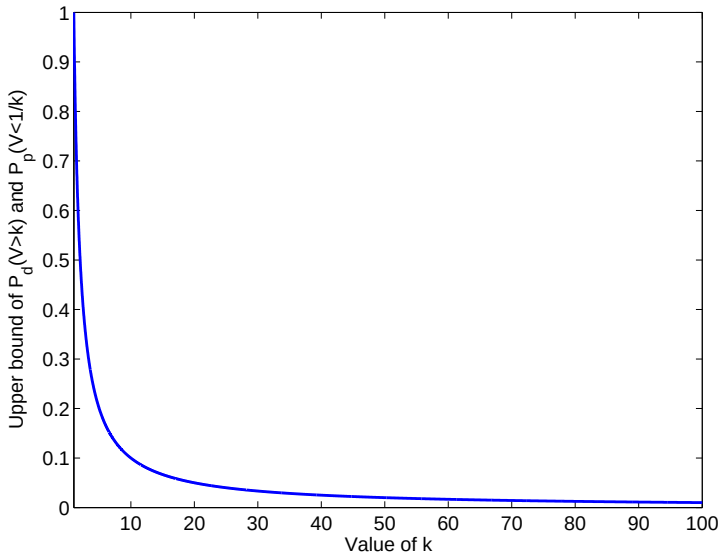


Figure 1: Universal bound on the probability of misleading evidence as a function of k>1.

8.3.2 Normal assumption, equal variances

Here we describe a simplified scenario where the distributions of the data are assumed to be normal with equal variances. This distributional assumption founds application in many domains in forensic sciences, particularly when biometric systems are used to analyse the evidence (Gonzalez et al., 2007, pp. 2104-2115; Ramos et al., 2008; Ramos et al., 2013a). Topics related with the theoretical behaviour of biometric systems with relation to normal distributions have attracted interest in recent years (Van Leeuwen et al., 2013; Brummer et al., 2014). Nevertheless, this assumption is theoretically tractable, and therefore it is useful in order to illustrate how tighter bounds than the universal bounds could be obtained in some cases, as it happens in (Royall, 2000, pp. 760-780).

Let us assume a sample X_1, X_2, …, X_n of independent and identically distributed random variables with normal hypothesis-dependent distributions $f_p = N(\mu_p, \sigma)$ and $f_d = N(\mu_d, \sigma)$. From (Royall, 2000, pp. 760-780), in this case the upper bound on the probability of strong misleading evidence is given by the so-called 'bump function':

$$P_d(V>k) = P_p\left(V < 1/k\right) < \Phi\left(-\frac{c}{2} - \frac{\log k}{c}\right) \tag{2}$$

where

- Φ denotes the standard normal cumulative density function.
- $c = \frac{\Delta\sqrt{n}}{\sigma}$ is the distance of the mean values, $\Delta = |\mu_1 - \mu_2|$, expressed in standard errors.

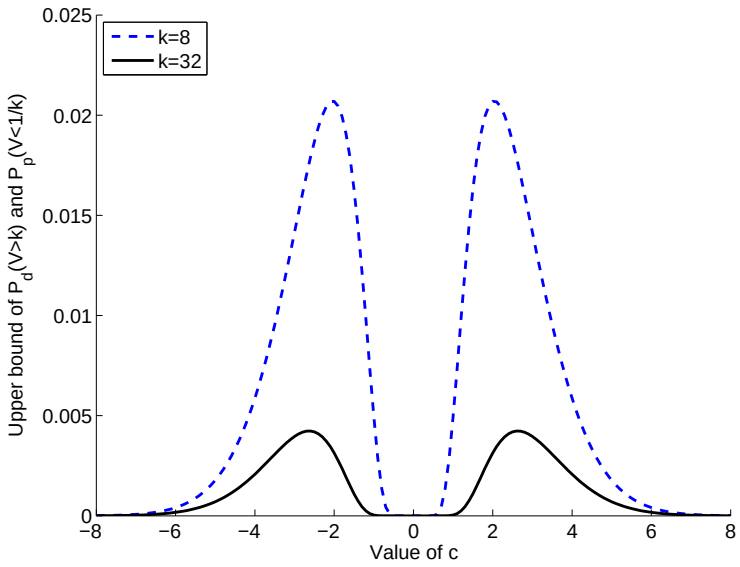


Figure 2: The bump function is the upper bound on the probability of strong misleading evidence for normal distributions with equal variances as a function of c, the difference of the means measured in standard errors.

The bump function is a function of c and k, and therefore it can be represented as a function of both variables, as it is shown in Figure 3. From such a representation, it can be seen that the bump function always decreases with k for fixed values of c. As in the case of the universal bound, this makes sense because increasing values of k decreases the cases where the value of V may exceed k. However, another effect is remarkable, because for different values of c the convexity of the bound of the probability of strong misleading evidence changes significantly. This is observed in Figure 4.

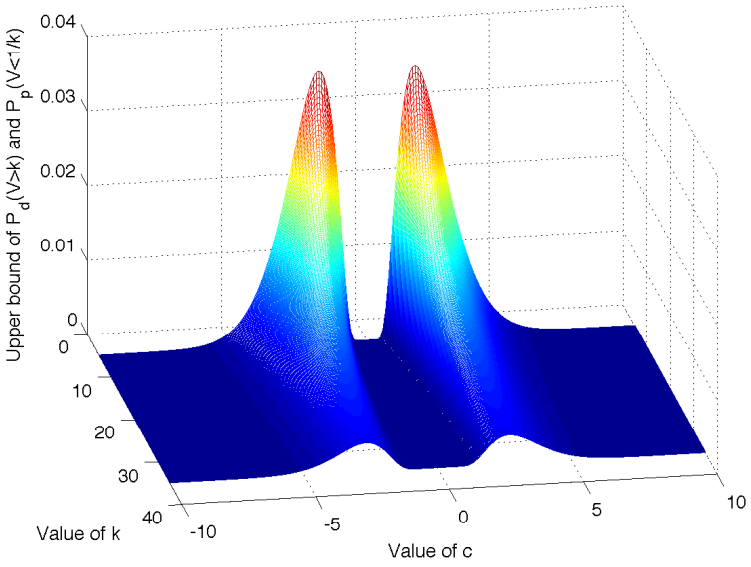


Figure 3: Bump function as a function of c and k.

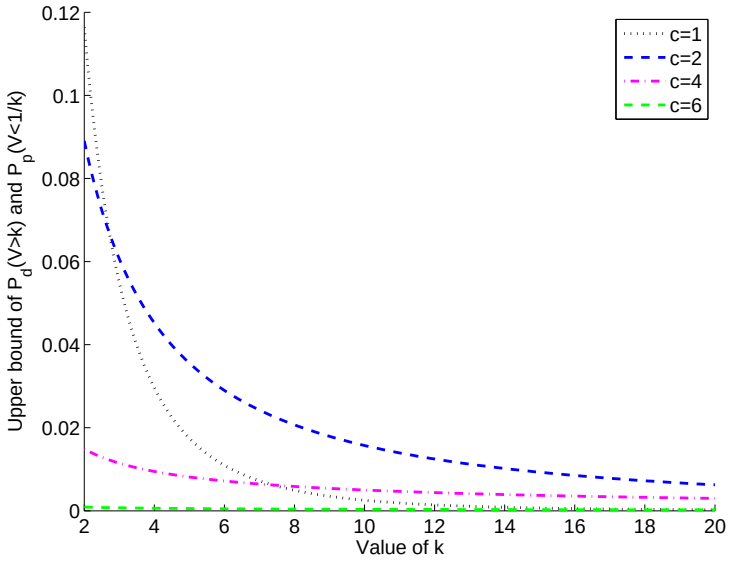


Figure 4: Upper bound on the probability of strong misleading evidence assuming normal distributions with equal variances. Represented as a function of k and for different values of c, the difference of the means measured in standard errors.

The maximum value of the bound on the probability of strong misleading evidence for a given k significantly varies with c, because the maximum value of the bump function is $\Phi\left(-\sqrt{2\log k}\right)$ for $c=\sqrt{2\log k}$. This result allows us to plot the 'maximum' upper bound on the probability of strong misleading evidence for 'any' pair of normal distributions with equal variances, regardless of the difference in their means and as a function of k. With this result, it is not needed to know the means of the hypothesis-dependent distributions in order to limit the probability of strong misleading evidence. Figure 5 shows the behaviour of such maximum upper bound with respect to the ones for different values of c as a function of k. We see that the difference among such bounds is sometimes quite important. The bound also has a decreasing behaviour with k, as expected.

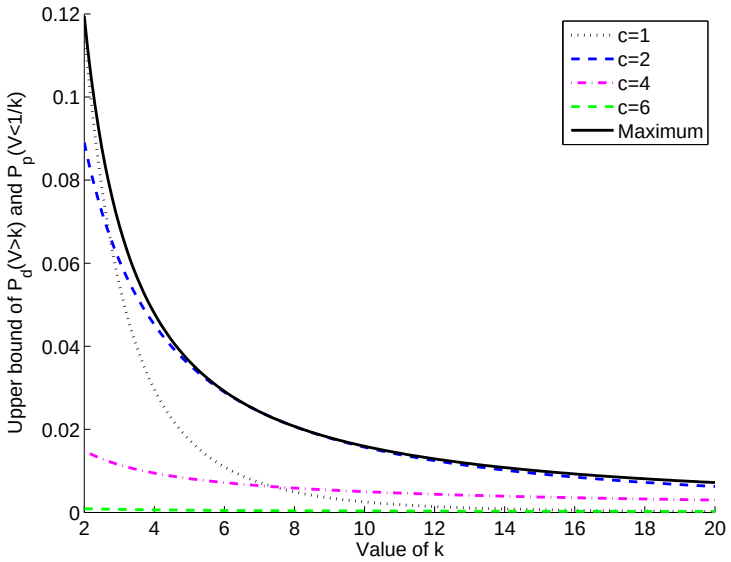


Figure 5: Maximum upper bound on the probability of strong misleading evidence assuming normal distributions with equal variances. Represented as a function of k and for different values of c, the difference of the means measured in standard errors.

The bound on the probability of strong misleading evidence for normal distributions with equal variances is much more restrictive than the universal bound. This is because a given distribution cannot have a looser bound than the universal bound for all distributions. Figure 6 shows the comparison between the universal bound and the normal bound under the equal-variance assumption, both for the maximum value of the normal bound, and for different values of c.

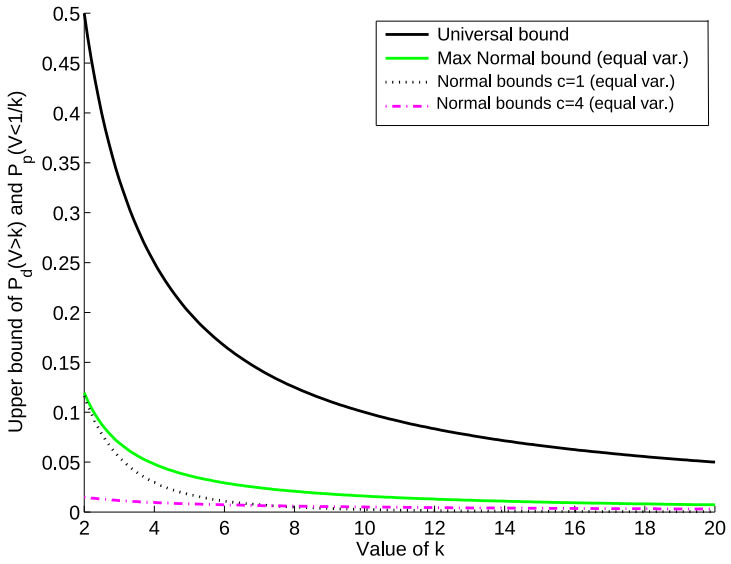


Figure 6: Comparison between the universal bound for the probability of misleading evidence and the bounds for normal distributions with equal variances.

8.3.3 Normal assumption, general case

To our knowledge, the solution for the bound on the probability of strong misleading evidence for normal distributions with different means and variances is not easily tractable using analytical methods. An example of this difficulty can be found in (Alcon, 2007). Also, a simulated approach from the input parameters by randomly generating samples in order to determine the probability of misleading evidence can be used in order to determine the bounds in this case. However, for the purpose and contribution of this article, this is outwith the scope of this work.

8.4 Application to the performance assessment of LR-based evidence evaluation methods

8.4.1 Empirical performance assessment of LR methods

The proposed methodology for performance assessment of LR-based evidence evaluation methods is empirical, i.e. it requires the availability of a 'validation' database that will be used for building simulated real cases (Ramos et al., 2013b). Thus, from those simulated cases, the performance can be measured. However, in other forensic disciplines, this kind of validation

database may not be available, or may be deemed non-representative, and therefore this empirical performance assessment methodology is not recommended. Anyway, data availability and representativeness for empirical performance evaluation of LR methods remains an open problem, which is outwith the scope of this work.

8.4.2 Tippett plots

Tippett plots (Evett et al., 1996b, pp. 79-86) are a valuable tool for the assessment of LR-based methods for evidence evaluation, empirically representing the cumulative proportions of the LR in an experimental set depending on which proposition was true (H_p or H_d). Figure 7 shows an example of Tippett plots. Important performance measures which can be seen in the plots are the rates of misleading evidence at the value of V = 1. Moreover, any rate of misleading evidence can be seen from Tippett plots for any value k > 1 or 1/k < 1.

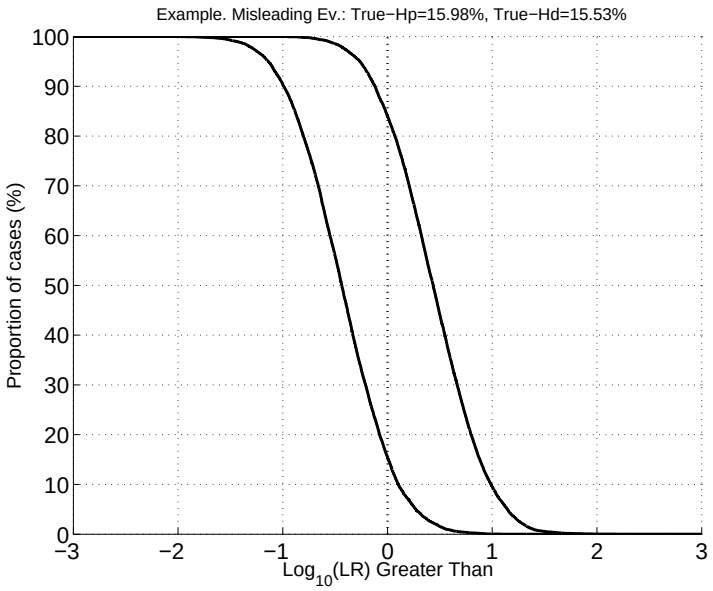
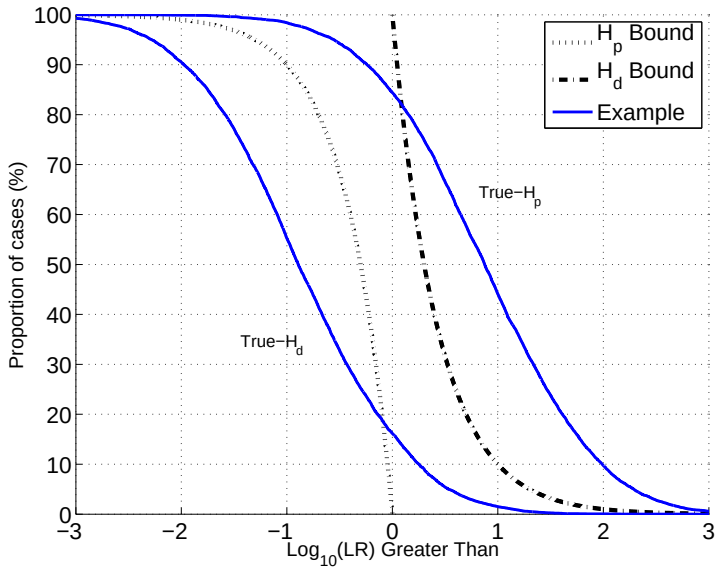


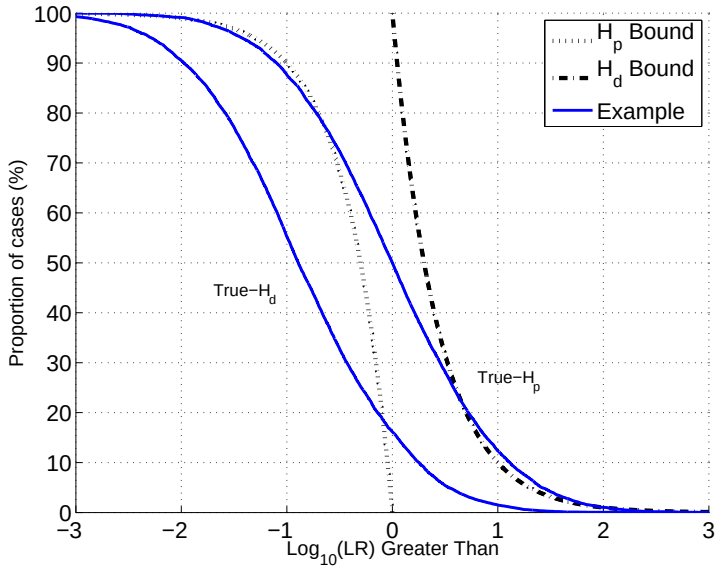
Figure 7: Example of Tippett plots. In the left curve H_d is considered true, and in the right curve H_p is considered true. The rates of misleading evidence at the value V = 1, represented in the title of the figure, are the proportion of cases where the value of the LR supports the wrong hypothesis.

8.4.3 Misleading evidence and Tippett plots: 'Limit Tippett Plots'

Tippett plots can be interpreted as representing the probability of observing a $\log_{10}$-LR value greater than a given value in the x axis. Such a number may be a given value of misleading evidence previously stated, say k. Therefore, we can represent the bounds derived in Section 8.3 for the probability of strong misleading evidence in Tippett plots. An example of that is shown in Figure 8, where it can be seen that in one of the cases the computed Tippett plots exceed the theoretical bounds.



(a)



(b)

Figure 8: 'Limit Tippett Plots', i.e. Tippett plots with the corresponding universal bounds on the probability of misleading evidence as a reference. In (a) both curves lay within the theoretical bounds. However, in (b) such bounds are exceeded when H_p is true.

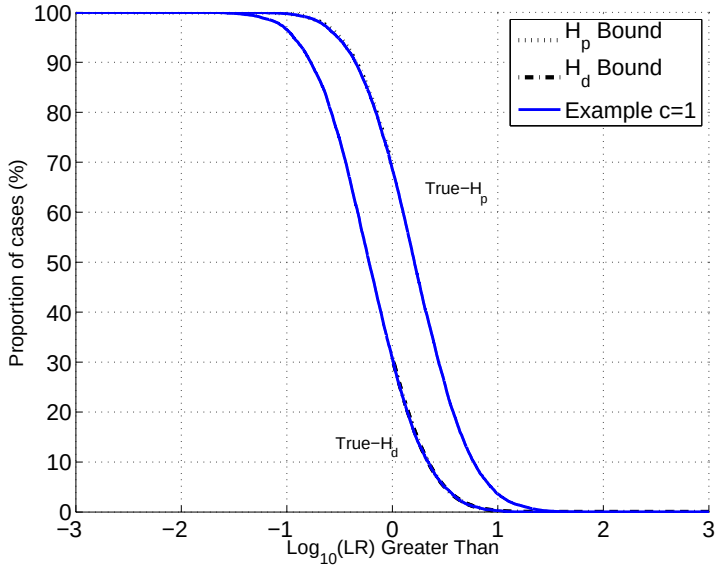
Note that the bounds when H_d is true are essentially the same that were derived in Section 8.3 (see e.g. Figure 1 for the universal bound). For the bounds when H_p is true, just note that, for the universal bound:

$$P_p(V < 1/k) = P_p(-\log V < \log k) = 1 - P_p(-\log V > \log k) \quad (3)$$

This allows plotting the bound when H_p is true by simply flipping the bound when H_d is true over log V = 0 and obtaining the complementary bound.

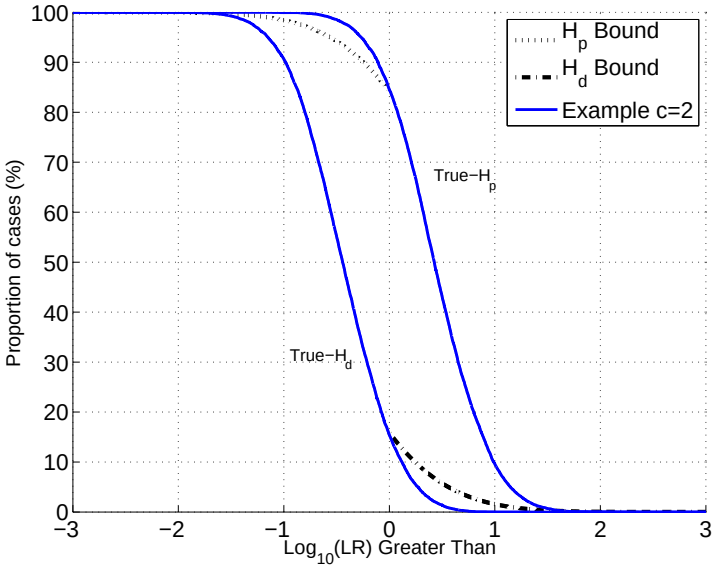
The bounds on the probability of strong misleading evidence for normal distributions with equal variances can also be drawn in Tippett plots. It still holds that such bounds are equal for $P_d(V > k)$ and $P_p(V < 1/k)$. Figure 9 shows several examples of Tippett plots including such bounds for different values of c, and for the maximum bound according to the bump function.

For Figures 9a-c, Tippett plots have been generated according to the distribution of each bound which is represented (e.g. a distribution having c = 2 has generated the example Tippett plots in Figure 9b). It can be observed that, again, the bounds on the probability of strong misleading evidence are much more restrictive for the normal distribution than the universal bound, as it was argued in previous discussion. Moreover, as k approaches the value 1, this difference increases. A special mention deserves the case for c = 1 (Figure 9a), where the example Tippett plot almost exactly reaches the theoretical bound. The maximum value of the bound, as expected, is looser than for any fixed-c bound. The bounds without their corresponding Tippett plots are represented in Figure 10 for illustration[22].

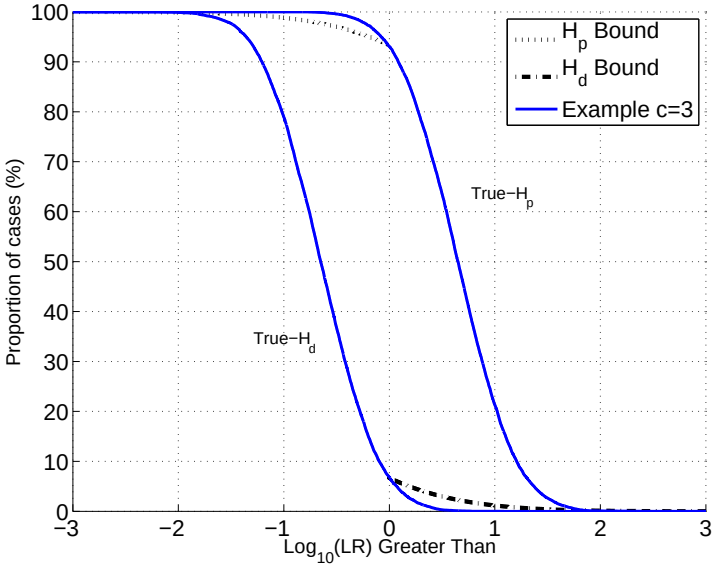


(a)

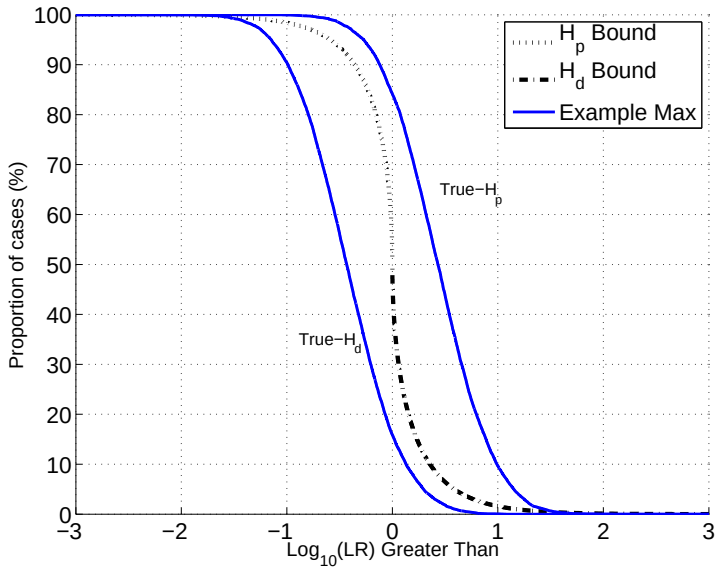
[22] Evidence is said to be misleading when the log(LR)>0 for H_d true, and when the log(LR)<0 for H_p true. According to the explanation above, both situations require to be considered separately, and that is the reason why the curves are different for log(LR) < 0 and log(LR) > 0 in the limit Tippett plots. Therefore, at log(LR) = 0, the requirements of the misleading evidence can be different when log(LR) approaches 0 from positive values (i.e. when H_p is true) and from negative values (i.e. when H_d is true). This explains the discontinuities at log(LR) = 0 for the curves for different values of c.


100
90
80
70
60
50
40
30
20
10
0
−3
−2
−1
0
1
2
3
Proportion of cases (%)
Log10(LR) Greater Than
Hp Bound
Hd Bound
Example c=2
True−Hp
True−Hd

(b)

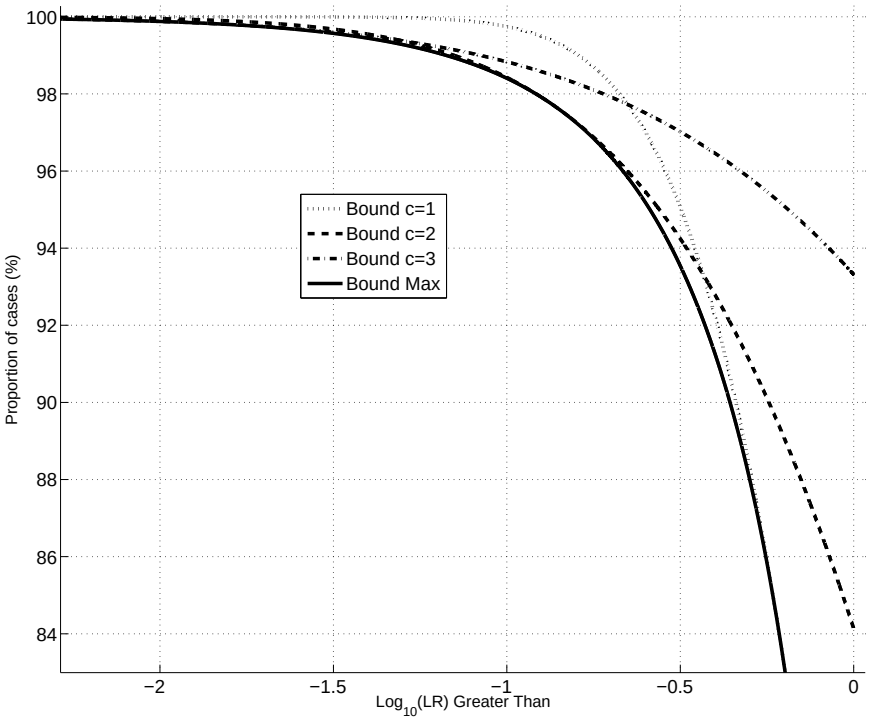

100
90
80
70
60
50
40
30
20
10
0
−3
−2
−1
0
1
2
3
Proportion of cases (%)
Log10(LR) Greater Than
Hp Bound
Hd Bound
Example c=3
True−Hp
True−Hd

(c)

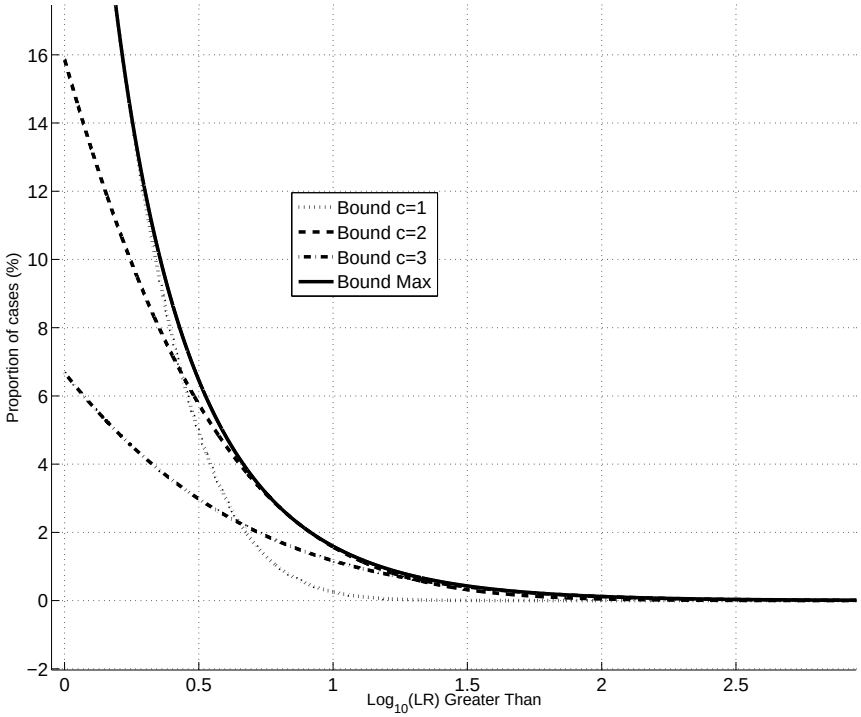


(d)

Figure 9: 'Limit Tippett Plots' with the corresponding bounds on the probability of misleading evidence as a reference, for the case of normal distributions with equal variances. Different values of c are represented, namely c = 1 (a), c = 2 (b) and c = 3 (c). The maximum bound with respect to c for all k is shown in (d).



(a)



(b)

Figure 10: Bounds on the probability of misleading evidence for the case of normal distributions with equal variances.

Tippett plots exceeding the theoretical bounds indicates that there are problems in the LR computation process, either e.g. because of the selection of the statistical models, or because the use of inappropriate data. Therefore, the use of 'Limit Tippett Plots' allows warning about those problems, which may not be seen in Tippett plots, and fostering the improvement of the LR computation methods themselves. Although the way in which the improvements of methods depends on many factors (databases, statistical models, type of data, etc.), detecting violations of the theoretical bounds can be applied to any method computing LR values, and therefore the usefulness of 'Limit Tippett Plots' is general for any LR-based discipline.

8.5 Experimental example: forensic automatic speaker recognition

In this section we describe an experimental example in order to illustrate the usefulness of 'Limit Tippett Plots' to detect problems in statistical LR computation methods. The example is contextualized as an experiment to measure performance of a forensic automatic speaker recognition following the procedures used in the AL-GC. These procedures have been recently accredited by the Spanish Accreditation Body (ENAC, www.enac.es).

It is worth noting that, although in this article we use the example of forensic automatic speaker recognition to illustrate the use of the proposed methodology, it can be applied to any forensic discipline where LRs are used as the expression of the strength of the evidence, and where an empirical performance assessment is in order. Therefore, the following experimental section can be understood without previous knowledge in forensic automatic speaker recognition. In the experimental example, we follow the procedures of data selection in forensic automatic speaker recognition in Guardia Civil, and then we assume that the validation database used in that example is representing different scenarios in casework, mainly defined by the conditions of the recordings (mainly related to the transmission channel of the speech signal). Therefore this empirical experiment can be seen as representative of future performance in casework where the conditions of the evidential recordings may fit those of the considered scenarios.

8.6 Context and Motivation

AL-GC regularly submits forensic reports to court based on a likelihood ratio methodology, for which an automatic forensic speaker recognition system is used. This system is described in (Kenny et al., 2008, pp. 980-988). LRs are computed by the system using procedures previously reported in (Gonzalez et al., 2007).

To compute LRs, a population of recordings from speakers is needed, and therefore a database must be used to select that population. The use of an appropriate population database is important in order to compute the LR. This case study shows how 'Limit Tippett Plots' are useful in order to detect violations of the universal bounds of misleading evidence, and therefore problems in likelihood ratios due to a bad selection of the population.

The selection of a population in LR-based evidence evaluation should consider the circumstances of the case and the propositions defined for the case. In this experimental example, there are also issues concerning the conditions of the recordings used in populations (e.g. noise level, transmission channel, speaker emotional state, etc.). The protocols in Guardia Civil are strict about those conditions, and therefore, the cases where the available populations are not appropriate for the conditions of the evidential recordings are typically rejected by the laboratory.

8.7 Forensic Scenario

The forensic scenario in which the case study is described represents typical conditions in AL-GC casework. In particular, two recordings of speech (one questioned and the other one coming from a given suspect) are to be compared, in order to yield a LR using the automatic system. The two recordings are known to come from different sessions, and from different moments in time. Both recordings are known to come from digital wiretaps of GSM mobile conversations, obtained in accordance to operational procedures in Guardia Civil.

8.8 Validation Database

To generate the LR values that will be represented in 'Limit Tippett Plots', a so-called 'validation database' will be used (Ramos et al., 2013b) simulating the conditions found in casework. The aim is to measure the performance of the method in use in the conditions in which it will be used in the case at hand. In the aforementioned forensic scenario, the 'Ahumada-IV' database is used by AL-GC as a validation database. The database consists of 91 male speakers speaking in different dialects of Spanish from Spain. The identity of the speaker is known, and the utterances contain spontaneous speech presenting high variability in recording times, acoustic environment (there are indoors and outdoors recordings), dialect, noise, etc. All the recordings are acquired using the wiretapping system used for obtaining the recordings in the described forensic scenario. For each speaker, a 'long' recording of roughly 120s and about 5 'short' recordings of 20s are available. All recordings are from different sessions.

The experimental protocol to measure performance consists of computing a single LR value for each comparison between each 'long' recording and each of the 20s 'short' recordings. For each comparison, a LR will be generated. A comparison is denoted a true-H_p comparison if the 'long' and the 'short' recordings come from the same speaker. Conversely, a comparison will be denoted a true-H_d comparison if the 'long' and the 'short' recordings come from different speakers. Also, the LR values will be called true-H_p LR values or true-H_d LR values depending on whether they are respectively computed for a true-H_p or a true-H_d comparison. Using this protocol, a total of 442 true-H_p LR values and a total of 39780 true-H_d LR values are obtained. These LR values will be used to draw 'Limit Tippett Plots'.

To illustrate the effect of the population selection in the performance of the LR values, two different experiments will be conducted, each one using a different population database.

- In the first experiment, namely Ahumada-IV-Population experiment, a population that presents the same conditions as the validation database is used, which is highly desirable in forensic automatic speaker recognition (Ramos, 2007). This is the population used in the usual protocol followed by AL-GC in real cases for this forensic scenario where all the speech in the case comes from digital wiretapping.

- In the second experiment, namely Ahumada-III-Population experiment, the population used to compute LR values presents different conditions than the validation database. In this case, it is expected that the models obtained from the population will not be so adequate for the data in the validation database, and therefore we expect worse performance in the Ahumada-III-Population experiment than in the Ahumada-IV-Population experiment.

8.9 Population Databases

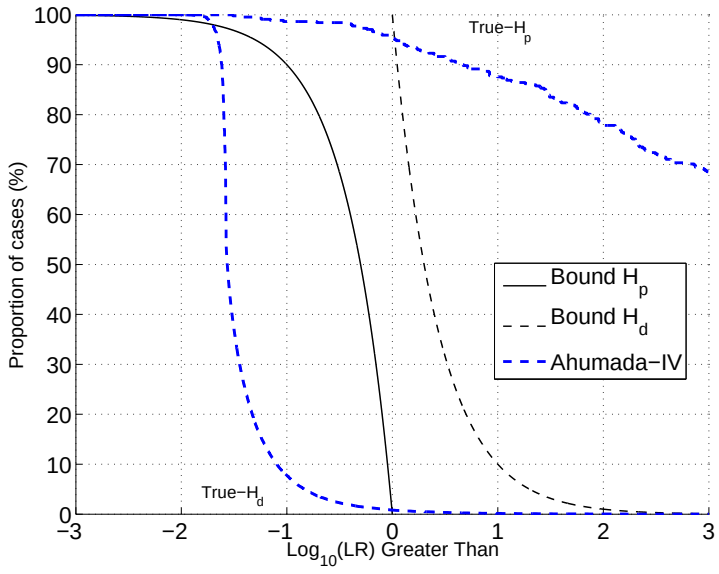
Here we describe the databases used as the reference population in both experiments presented.

- Ahumada-IV-Population: this population database comes from recordings from the Ahumada-IV database itself. From all the 'long' recordings in the database, a total of 35 recordings are used as the population. In each comparison, the population selected did not include the recording corresponding to any of the speakers in the comparison. This population presents the same conditions as the validation database, because it is actually extracted from the same database, but using recordings from different speakers than the ones in the case. Therefore, it consists of recordings of 120s, obtained using digital wiretapping.
- Ahumada-III-Population: this population database comes from the so-called Ahumada-III database, described in (Ramos et al., 2008, pp. 1493-1496). The conditions of this database are different from the conditions in Ahumada-IV. In Ahumada-III, the recordings also originate from GSM mobile conversations, but they are recorded over magnetic tape, not digitally. This emulates the procedures used by AL-GC before 2005, year in which the digital wiretapping system started operation in Spain. Ahumada-III is a database of real forensic cases, and the Spanish law has allowed AL-GC to use it for forensic purposes. For the Ahumada-III-population experiment presented here, 69 recordings of about 120s have been used as the population. However, although the speech in Ahumada-III database comes from real forensic cases, there is a substantial difference between the conditions in the Ahumada-III and Ahumada-IV databases, because the recording technique used in both databases is also different. This suggests that the performance in the Ahumada-III-population experiment will be worse than in the Ahumada-IV-population experiment.

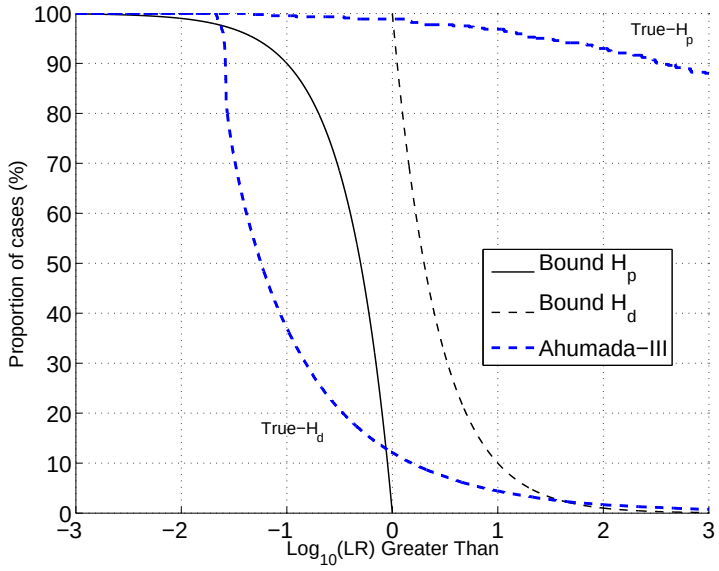
8.10 Results

In this section, we present results that show the usefulness of 'Limit Tippett Plots' as tools for measuring performance.

Figure 11a shows the 'Limit Tippett Plots' of the Ahumada-IV-Population experiment. As it can be seen, the Tippett plots do not violate the theoretical universal bounds for the probability of misleading evidence. According to (Ramos, 2007), this can be partially because a proper population has been used to model between-source variability, since the conditions of the population are the same as the conditions of the cases in the validation database. Remarkably, with 'Limit Tippett Plots' the adequacy of the LR values to the theoretical bounds of misleading evidence is made explicit, and consequently it adds valuable information to regular Tippett plots.



(a)



(b)

Figure 11: 'Limit Tippett Plots' for the Ahumada-IV-population experiment (a) and for the Ahumada-III-population experiment (b).

Conversely, Figure 11b shows the 'Limit Tippett Plots' of the Ahumada-III-Population experiment. As it can be seen, in this case the Tippett plots violate the theoretical universal bound for the probability of misleading evidence when H_d is true: it can be seen that the proportion of true-H_d LR values greater than $10^{1.6}$ is greater than $10^{-1.6}$, and therefore the bound is exceeded. This can be attributed to a bad selection of the relevant population for LR computation, because the conditions of the Ahumada-III-population (GSM recorded over magnetic tape) are very different than the conditions of the validation database, Ahumada-IV

(digital GSM wiretaps). Moreover, the Ahumada-IV-population experiment and the Ahumada-III-population experiment only differ in the population in use, and therefore the loss of performance can be attributed to this fact. Therefore, the violation of the universal bounds constitutes a warning about the adequacy of LR computation methods, in this case because the selection of the population was inadequate. This information can be seen in 'Limit Tippett Plots' because of the explicit representation of the universal bound, but not in regular Tippett plots. In order to better illustrate the aforementioned observations, a detailed comparison of the 'Limit Tippett Plots' for both experiments is shown in Figure 12, where the region of the plots where the universal bound is violated has been magnified.

Sample size effects may have an influence in the violation of the theoretical bounds in 'Limit Tippett Plots'. In an experimental setup, the smaller the size of the validation database, the bigger the expected variability in the figures of performance (e.g. rates of misleading evidence in Tippett plots). As the database increases in size, those figures will be more robust, and their credibility will be better. This also applies to the exact point where the theoretical bounds of misleading evidence are violated. Therefore, an increase in the size of the database is always desirable. In the current example, speech databases are always very costly to be acquired. Although there are initiatives such as the NIST Speaker Recognition Evaluation campaigns[23], the protocols of any forensic laboratory need to have an available database that is as much similar to the casework as possible, and this is not always an easy task. This is why studies of the influence of data sparsity and other effects are extremely necessary in many fields (Tapias, 2005), as well as constant data collection in forensic conditions.

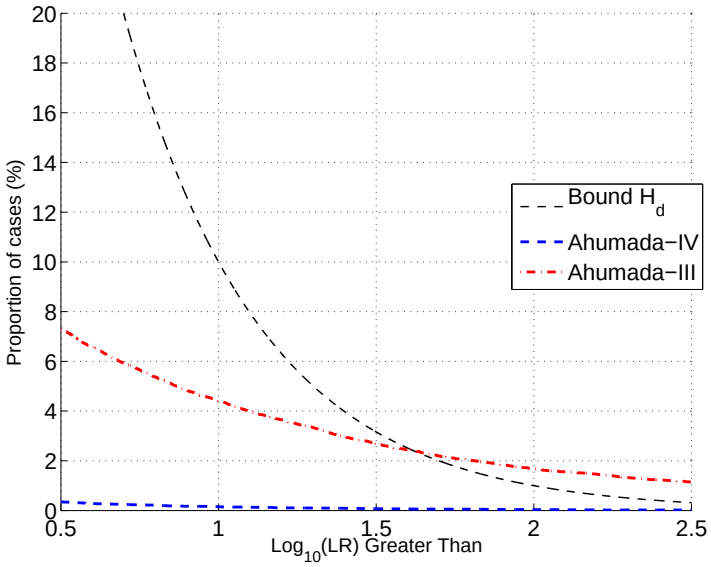


Figure 12: comparison of the 'Limit Tippett Plots' for the Ahumada-IV-population and the Ahumada-III-population experiments, where the region where the universal bound for misleading evidence is exceeded has been magnified.

8.11 Conclusions

This work contributes with a novel tool to detect likelihood ratios presenting too high misleading evidence, by the proposal of so-called 'Limit Tippett Plots'. Based on previous

[23] See https://www.nist.gov/itl/iad/mig/speaker-recognition-evaluation-2016 (accessed on January 12, 2017).

contributions (Royall, 1997; Royall, 2000, pp. 760-780; Aitken and Taroni, 2004), the behaviour of the bounds of the probability of observing strong misleading evidence in a set of LR values has been described. Two main conditions have been examined: unknown data distributions (universal bounds), and normal data distributions. Then, Tippett plots including the theoretical bounds of the probability of misleading evidence, namely 'Limit Tippett Plots', have been proposed for detecting an anomalous behaviour in LR values. This proposal is useful for examiners conducting experiments with the aim of measuring performance of LR values. A Matlab™ software has been made available by the authors, in order to easily draw 'Limit Tippett Plots'. It can be downloaded at http://arantxa.ii.uam.es/~dramos/software.html.

Finally, an experimental example in forensic automatic speaker recognition has been presented. There, the protocols followed by the AL-GC are followed in order to measure performance in different scenarios, and 'Limit Tippett Plots' are used in order to detect problems due to an inadequate selection of the population used to produce LR values.

Chapter 9. EPISTEMOLOGY APPLIED TO CONCLUSIONS OF EXPERT REPORTS[24]

Abstract

It is believed that to build a robust reasoning logic to make probabilistic inferences in forensic science from a merely mathematical or logistical viewpoint is not enough. Mathematical logic is the positive science of reasoning and as for that it is only interested in the positive calculus of its validity, regardless any prior ontological assumption. But without a determined ontology and epistemology which imply to define the concepts that they will use, it seems difficult that the proposed scientifically correct mathematical solution be successful as a European standard for making conclusions in forensic reports because it has to be based on judicial language.

Forensic experts and Courts are not interested in the development of a positive science but in a practical science: in clarifying whether certain known facts are related to a possible crime. Therefore, not only the coherence of the demonstrative logic reasoning used (logic of propositions) is important, but also the precision of the concepts used by language and consistency among them in reasoning (logic of concepts).

There is a linguistic level essential for a successful communication between the forensic practitioner and the Court which is mainly related, in our opinion, to semantics and figures of speech. The first one is involved because words used in forensic conclusions often have different meanings – it is said that they are polysemic - and the second one because there is often metonymy as well. Besides, semantic differences among languages regarding words with the same etymological root add another difficulty for a better mutual understanding.

The two main European judicial systems inherit a wide and deep culture related to evidence in criminal proceedings and each of them has coined their own terminology but there are other two more abstract levels such as logical and epistemological, where we can find solid arguments by which terms used at legal level on conclusions of forensic reports could be accurate and consistent for all users of an intended EU guideline. An effort has been made to elucidate the following terms: truth, certainty, uncertainty, opinion, conjecture, probability, evidence, belief, credibility, determinism, indeterminacy, cause, principle, condition, and occasion.

Keywords: conclusions; truth; uncertainty; evidence; belief; credibility.

9.1 Introduction

The development of a specific technical guide on evaluative forensic conclusions (ENFSI guideline, 2015) has emphasized the need for using a glossary of terms acceptable both by two very different judicial systems in Europe, the Continental and the Anglo-Saxon ones, and by the special characteristics of the coexistence of so many languages in it. Terminological differences based on philosophical viewpoints were noted during the development of the project. These differences justify important semantic changes in words of current languages derived from the same Latin or Greek words as evidence, uncertainty, doubt, certainty, truth, probability or

[24] LUCENA MOLINA JJ, "Epistemology applied to conclusions of expert reports". Forensic Science International, Special Issue EAFS 2015, 264, July 2016, pp. 122-131.

proof, which have an unquestionable importance when defending an expert report before the Court[25].

If many forensic experts have emphasized the relevance of a closer scene of crime inspection they should spare no expense with the expert report conclusions. Undoubtedly, this last assertion is still being a target without our reach in Europe, although the appearance of the publication of a technical guide on evaluative conclusions lays the foundations for a paradigm shift in forensic science that some authors have described as the most important consequence of the advances made by that science in the 20th century (Evett, 2009, pp. 159-160).

According to general legal doctrine about criteria for interpretation of legal texts, the grammatical meaning of legal texts is the starting point to interpret any norm. However, it is also important to bear in mind the context in which those terms are used, which is known as the logical or systematic criterion. Both criteria, grammatical and logical, are mutually involved in our case. Therefore, legal terms used in laws of criminal proceedings or analogue regulations need to be understood by procedural actors respecting those criteria unequivocally.

Philosophical considerations are present in Bayesian literature applied to Forensic Science. One specific theory of knowledge is behind some key terms needed for describing, for example, the concept of subjective probability as a degree of belief (Taroni et al., 2014). Besides, judicial language applied in the context of conclusions of forensic reports entails to write or speak about truth, belief, certainty, uncertainty, doubt, consistency, causes and effects, explanations, and so on. It is not easy to find deep explanations of those concepts in forensic scientific literature because, in fact, it has mainly been a duty of philosophers for centuries.

In the author's opinion, some especially relevant concepts used in the works of the main researches which defend the likelihood principle in forensic science are influenced by the logical positivism. The current work studies some of these terms, used in those writings with a strongly biased semantic positivism, and compares their meanings with those used in the Aristotle-Thomas school. It is intended to show that for this philosophical school the subjective concept of probability[26] is perfectly assumable. However, it is pointed out that the confusion between the verbs 'to know' and 'to believe' that the author of this work perceives in the

[25] Though (R. Allen, 2015, p. 71) states that: "The standard critique assumes or asserts that an important goal of the legal system is to admit or take advantage of scientific knowledge, and then descends into the seemingly endless (to the not-so-philosophically minded) wrangling over what is knowledge and how we can know that we possess it", in our view, this comment is a consequence of the predominant 20th century Anglo-American theory of knowledge in the Anglo-American legal system (see Sections 9.2.1 and 9.2.2). Therefore, we think that R. Allen deals with a specific problem of the Anglo-American legal world. The philosophical discussion on the true concepts of knowledge and belief, and the application of the Anglo-American epistemology within the criminal proceedings of Continental legal systems simply does not exist (see Lucena et al., 2012).

[26] A. Biedermann (Biedermann, 2015, p. 142) distinguishes between a formalist and an intuitionist-constructivist view of mathematics. Focusing his attention on the theory of probability, the former could be characterised saying that "the meaning assigned to the terms that make up the formulation of the laws of probability could be considered independently from the formulation of the laws themselves", and the latter as "a formal language for saying specific things". A. Biedermann underlines: "Every formula in the language means something, beginning with the most basic items. In particular, in de Finetti's construction 'the laws' derive from meaningful assertions regarding prices."

positivist justification theory implies an intrinsic difficulty to understand and undoubtedly accept the concept of subjective probability.

Four different but complementary epistemological fields are harmonised in this work: ordinary knowledge, law, philosophy and science. They are indeed irreplaceable fields. It is believed that this interdisciplinary approach allows researcher to study the evaluative reporting in Forensic Science from an outstanding viewpoint.

The purpose of the paper here is to discuss what some scholars have proposed in the past as guidance and how we can distinguish between moderate realism and immanentism for clarification purposes briefly (Section 9.2); to point out some relevant aspects from the philosophy of law to fix the context of its title (Section 9.3); to apply epistemology to specific terms related to conclusions of expert reports (Section 9.4); and finally, to draw some conclusions (Section 9.5).

9.2 Historical background

9.2.1 Epistemology in the 20th century

Some outstanding philosophers of law have studied the epistemological model of judicial determination of facts according to the evolution of the predominant theories of knowledge in each historical stage. Gascon (Gascon, 2010, pp. 45-67) points out the following predominant theories of truth in the 20th century: the coherentist theory, defended by Bradley, Neurath, Rescher and Dauer; the pragmatist theory proposed, with different shades, by Pierce, James, Dewey, Kuhn, Habermas and Putnam; and the correspondence theory advocated by Taruffo, Ferrajoli and Tarski, among others.

Gascon chooses the correspondence theory for the very reason that it is the unique among the three ones that, in her opinion, complies with the aim of the judicial process to investigate criminal facts. She states that "both, in common and judicial languages it is not asserted that a set of statements on facts is true because it is internally coherent – then, a novel would be it -, because it is accepted – even unanimously -, or because it is simpler than others, but the described facts have really happened." And she continues further to say the following: "… the coherentist and pragmatist concepts disconnect from the aim of the judicial process to set facts, that it is the very reconstruction of facts, as those happened." Gascon also states that the correspondence theory "is based on a minimum realist philosophy that, without despising the theoretical conditioning, and, above all things, regarding the judicial-institutional process 'to access to the facts' - , allow us to keep the hope of having an objective knowledge." And she finally adds that "The majority of the realist epistemologies bet on the realism precisely to justify the objectivity of knowledge."

However, the coherentist and pragmatist theories of truth were predominant in the second half of the 20th century Anglo American epistemology (Moros et al., 2003, pp. 633-671) and it can be appreciated its influence in the Bayesian literature.

The concept of belief has been used by Anglo American epistemologists as an essential element to define the concept of knowledge. They linked it to the Greek word 'doxa' which may be translated as opinion. Therefore, for them, belief and opinion mean, in general terms, the same thing.

The term ‘belief’ shows the following features (Vernaux, 1971, p. 138): (i) it can comprise any kind of judgement – in its broadest sense – and then it is identified with what is consented, asserted, i.e. the judgement itself; (ii) it can also be an affirmation mixed with uncertainty, matching up with what we called opinion; and (iii) it can appoint to certainty in opposition to science: act of faith.

Therefore, one of the possible meanings of ‘belief’, in particular the one that identifies the verb ‘to believe’ with the verb ‘to know’ and which is expressed with the term ‘opinion’, is used by Anglo American epistemologists.

The quoted specific concept of belief is understood as a state of mind. Sometimes it is described either as inferior to knowledge (understood as objective knowledge), or a state of mind, by default, when a human being is willing to know something (finding objectivity). Therefore, it is a state of mind which we have to undergo necessarily if we want to reach objective truths.

For immanentists ‘to know’ has the same meaning as ‘to be willing to know’ because they deny the immediacy of knowledge. They deny the existence of the agent intellect (Corazon, 2002, pp. 75-79).

However, the term ‘belief’ is used in Anglo American epistemology with a more precise meaning than the term ‘opinion’. A judgement is called opinion when there is uncertainty in the mind of a person who knows and makes a judgement, that is, when there is fear to err. However, belief does not imply to assent the judgement necessarily for Anglo American epistemologists. Belief is understood as a disposition or habit of mind to accept something as true or probably/possibly true.

It might be understood before the vagueness of such a definition of ‘belief’ that the mentioned epistemologists make specific references in their supported definitions of knowledge to the fact that belief – assumed as true – has to be justified to speak about objective knowledge properly.

If justification turns a belief assumed to be true into knowledge, the debate is focused on the meaning of the term ‘justification’.

Since C. I. Lewis devised the concept of knowledge based on the above mentioned meaning of the term belief, subsequent developments to his thought were focused on the relationship between knowledge and truth, being the concept of justification the key to the conception of different theories:

a) Internalism: it underlines characteristics to which he/she who knows has access through his/her conscience. The following may be distinguished within internalists:

    - Foundationalists: they defend the existence of self-justifying beliefs to avoid an infinite regression.
    - Coherentists: they deny the existence of auto justified beliefs and underline that justification is based on the coherence among beliefs.

b) Externalism: it underlines characteristics external to subjects. Reliabilists have reached special importance among externalists, and they are mainly worried about the technical quality of cognitive processes.

The problem is the current lack of agreement among epistemologists of belief on what is exactly the term 'justification', though there are agreements on how they have to be reached, or how this belief has to be acknowledged once reached.

In any case, the quoted meanings of such terms are very far from those conventionally used in the judicial scope. Moreover, the use of a vocabulary indistinct from that of classical epistemology as it happens with the terms 'truth', 'justification', 'belief', causes a phenomenon of polysemy which damages, in our view, the mutual understanding among procedural actors in courts[27].

9.2.2 Historical philosophical rivalry: moderate realism versus immanentism (Metaphysics, book X; Vernaux, 1971; Alvira, 2001; Corazon, 2002; Millan, 2002; Llano, 2007)

Aristotle's metaphysics of knowledge stated that evidence is the cause of certainty in the mind, and evidence is a property of the object. According to the school of thought born from Aristotle and called moderate realism, the process of understanding allows human beings to grasp the essence of things by the so-called simple apprehension, that is, to actually know what things are using their senses and the intelligence; to make statements or judgements dividing or composing concepts; to link the quality of true to those statements after comparing them with reality, and to reason drawing conclusions from premises according to logical principles. What is known by the mind is called object, and it is not a synonym of thing, but something formal, universal, and immaterial.

For this school of thought, to know is to own an object. This ownership is intentional, that is, what is owned by the intelligence is the object. To know is a perfect act and, therefore, it reaches its aim directly and immediately. For example, to walk, to build, or to learn are examples of movements, but to see or to think are perfect acts because they reach their aims immediately as Aristotle states, e.g. we see and we have seen, we think and we have thought.

Summarizing, truth is based on existence, and evidence causes certainty in the mind.

Immanentists are nominalists, that is, they consider ideas to be mental fictions, useful to classify objects, but useless to reach true knowledge. They are common names to call things, and they lack intelligible content. Descartes equaled evidence and certainty. He lost the intentionality of knowledge focusing his attention on the object and questioned that it represents something real. He replaced intentionality by causality. Ideas are caused, and they refer to reality as cause, not as object. Therefore, from Descartes on, to know is not to open

[27] For example, the difficulties of comprehension of the Bayesian approach in DNA comparisons in the judicial scope are not only something related to epistemic roots of some concepts (subjective probability, likelihood ratio, and so on), but also to ontological roots of those concepts which they are based on (belief, degree of belief, uncertainty, and so on). Henri Poincaré is often quoted as the first scientist who used the Bayes' theorem in a forensic case and, obviously, he did not need the concepts developed by the above-mentioned predominant 20th century Anglo-American epistemology to make probabilistic inference in that case.

oneself to reality, to accept reality as it is, but to control reality. Before accepting any object as certain, it has to be exhaustively and previously examined to be accepted as clear and distinct by the mind.

Empiricism is nominalism. They also identify 'to know' with 'to be willing to know'. For them all true knowledge should be based on empirical verification. Existence is then understood as a fact, not as an act. It is the verification of the fulfillment of a hypothesis in a specific case. It is well-known that the end of this way of thinking has been the formulation of a subjective and utilitarian concept of truth, because truth is not already based on being but on essence, and there are real and thought essences. Therefore, if we do not base truth on being, or in existence, which is the effect of being, we will not be able to distinguish reality from thinking. Subjectivism is, in the last analysis, voluntarism. The subjective attitudes set the will before intelligence: they do not base certainty on evidence but vice versa, they consider what the subject admits as certain to be evident.

Therefore, truth is a product of reason, and certainty causes evidence and truth in the mind.

Such an intended attitude is not real because it is not possible to anticipate knowledge, to reach an earlier knowledge to all knowledge. To reflect, to think about thinking is always something posterior, because what is immediate is to know something. Our acts of understanding go before every voluntary act.

There is no room to philosophically explain in depth the differences among the concepts of truth, certainty, and evidence between the moderate realism and the immanentism, but the earlier explanations could be enough to underline that the actual difficulty to write a common glossary of terms related to the theory of knowledge is not a consequence of the existence of two main legal systems in Europe: the Continental and the Common ones, but the old, wide, and deep philosophical controversy about the theory of knowledge in our continent.

9.3 General epistemology foundations to prove facts

9.3.1 The proof of facts in criminal proceedings

Criminalistic laboratories usually assist Judicial Authorities preparing the expert reports needed to solve possible crimes. Also they help fact finders to focus their efforts on those hypotheses which appear more credible or, on the contrary, to rule them out. Data provided by experts to the criminal proceedings and their interpretation given the propositions supported by parties are the basis of the information that should be expected by the applicants of expert reports.

The moment an expert report reaches the highest legal-procedural relevance level is when it is considered as evidence, usually during the oral hearing, and respecting the principles of immediacy, contradiction, orality and publicity required by the procedural doctrine and case law to safeguard basic rights. Besides that evaluative process, and consistently with it, motivation and proved fact statements in sentence can be found. The rationality of these processes is linked to the rationality in which an expert bases the conclusions of his/her report and how he/she gives testimony of them on trial.

Therefore, the proof of facts – in which expert report is included – in a criminal proceeding is the legal-procedural framework to be examined from epistemology.

A preliminary approach to the issue at hand is the linguistic one used to describe and explain it. The term 'proof' makes up the starting point and legal epistemology experts emphasize its polysemic character, both in legal doctrine and case law. Moreover, they state the existence of a 'huge terminological chaos' and inconsistency when it is used (Gascon, p. 76, fn. 98).

Though an effort to clarify and specify the meaning of proving something is our goal, it is important to be aware that it is not the facts – by themselves – what can be proven but statements on facts. Facts happen and they can only be confirmed while they are happening. Once occurred they can be verified, but only statements on facts can really be verified.

The statements about the facts are judgements, assertions, affirmations, assents, propositions. To verify a statement consists on checking its truthfulness or falseness. Here we find another term needed for a precise definition due to its basic character: the concept of 'truth'. This concept is an unavoidable point of reference to define 'certainty'.

The above-mentioned chain of concepts, from proof to certainty, makes up a top priority epistemological framework to understand well the meaning of uncertainty linked to any scientific expert report.

It is believed that a merely phenomenological explanation of uncertainty is not enough to understand it properly, that is, showing that it is a fact that all human being should unavoidably and frequently live with: any future event is uncertain, a lot of past events as well, and quite a few events lived as current ones are full of uncertainty, either by ignorance, or by lack of necessary information.

General epistemology or knowledge criticism is a key element to go into the problem that we are trying to tackle with a solid rational argument: how to prove facts. And within that discipline the concept of certainty is classically dealt with.

9.3.2 The role of logic of propositions (Gascon, 2010, pp. 92-104)

We focus the discussion on the so-called indirect or circumstantial evidence, which imply the Judge to reconstruct a hypothesis about facts which may explain the evidential statements achieved.

The appropriate methodology of indirect evidence of facts is induction. However, it is convenient to distinguish, at least, two main characterizations of induction: 'in strictu sensu' and 'in lato sensu'.

The former is the most general and well-known concept, the reasoning process from particular instances to wider generalizations. The latter, however, is the reasoning where premises, even being true, do not provide enough foundation to support the truth of their conclusion, but only a probability may be reported. While in deductive arguments the progress from premises to conclusion is analytic (necessary), in inductive arguments it is synthetic (not necessary).

Therefore, due to the unnecessary character of inference, induction is defined as the reasoning process by which a fact which has already happened is explained by the supposed occurrence of non-observed or non-observable facts (i.e. facts that happened in the past without

witnesses) but when there is still some basis to infer the former. This type of induction is called 'reconstructive induction'.

Gascon points out that this is how indirect evidence of facts works, because 'to discover' facts related to the case at hand the Judge tries to provide an explanatory version of the existing evidence: from the existing evidence (p) an explanatory fact of the facts of the case (h) is concluded, which is an explanatory hypothesis of (p).

The non-observed explanatory fact (h) which is the conclusion of the indirect evidence proceeding explains the verified facts (p) because there is a link between h and p, so that if h were true, p would make sense (or it would be understandable). The schema is the following:

$$\begin{array}{l} h \rightarrow p \\ h \\ \hline \quad p \end{array}$$

However, the schema which starting from evidence (p) and the links among them and the hypothesis (h $\rightarrow$ p), concludes with the hypothesis h, is the following:

$$\begin{array}{l} h \rightarrow p \\ \quad p \\ \hline h \end{array}$$

The earlier schema, by itself, does not justify h as the conclusion of a deductive reasoning. If so, it would incur in the fallacy of consequent. Due to the probabilistic character of the inference, to the relationship between premises and conclusion, this is a clear example of inductive reasoning.

It is also possible to imagine a different schema where the major premise is an empirical law considered to be true according to past experiences (p $\rightarrow$ h). This does not change the inductive nature of the reasoning because those laws do not state that h is a necessary condition of p (i.e. that h is the unique possible explanation of p). They only state that h may be a possible explanation of p. They only state that the occurrence of h is probable, provided that p also happens. We show the following schema:

$$\begin{array}{l} p \rightarrow h \text{ is probable} \\ p \\ \hline h \end{array} \quad [\text{it is probable}]$$

Uncertainty of conclusions of syllogistic arguments cannot be inferior to the lesser of their premises. If the major premise is probabilistic, conclusion can only be probabilistic. A syllogism as such is inductive, not by means of how premises are related to their conclusion but by the logic framework of one of the premises.

C. Peirce and N.H. Hanson coined the terms 'abduction' and 'retroduction' to designate the probabilistic inference consisting of looking for an explanatory hypothesis of certain facts. It

can be described as a backward reasoning, as Sherlock Holmes did, different from deduction and induction.

Abduction is like an indirect evidence of facts characterised as a way of induction, the so-called induction 'in lato sensu'. It is a way of reasoning for making sensible conjectures about facts, starting from a rule and some certain facts: a surprising fact p is observed; if h were certain, it would be an explanation of p; therefore, there are reasons to support the truth of h. So, starting from some facts (p) and a rule whose consequent are those facts ($h \rightarrow p$), the facts that made up the antecedent of the rule (h) are reconstructed, and if they were true, they would explain why p occurred.

The schema of abduction suggested by Peirce is the following:

$$\begin{array}{l} h \rightarrow p \\ \quad\ p \\ \hline h \end{array}$$

This is the same schema than that shown for induction 'in lato sensu', capable of the above-mentioned change:

$$\begin{array}{l} p \rightarrow h \text{ is probable} \\ p \\ \hline h \end{array} \quad \text{[it is probable]}$$

Therefore, there is no difference between induction and abduction. They are inductive arguments in both cases, because though the premises are true, the conclusion will be only a conjecture.

Gascon concludes her argument by saying that from some evidential statements and certain causal regular rules (usually simple lessons of experience) a hypothesis about facts is reconstructed to explain evidence better than any other. As those causal regular rules are probabilistic and, besides, some evidence has been mediately obtained, the explanatory hypothesis of those facts can only be probable. Therefore, the representation of fact of a sentence is not sure. In strictu sensu, it can only be stated that the indirect evidence concludes by making an assumption, a statement that we consider to be true, but we do not know if it is true for certain.

So, Gascon recommends the following epistemological rules that bearing the fallible nature of indirectly proved explanatory facts in mind, those statements can be as close as possible to the truth:

a) To assert the truth of an explanatory fact, evidence is needed to prove it, either direct or deductive or indirect.
b) Anything allowing us to give relevant information about facts under investigation should be available for use.
c) A legal evaluation of evidence is rejected in favour of a free evaluation of evidence by the court as a methodological principle (negative) allowing it not to consider explanatory facts to be proved while the court does not agree with it.

d) There is never enough evidence. Any relevant evidence is necessary and, therefore, it should be admitted.
e) All parties should be offered the possibility to refute hypotheses (requirement of contradiction).
f) If in light of new evidence the truthfulness or falseness of a statement had to be revised, it should be done.

9.4 Discussion

9.4.1 Certainty and uncertainty

It does not make any sense to confuse two terms in the common spoken language whose meanings are, at first sight, contradictory. It would be a flagrant violation of the most basic principle of reasoning known as the non-contradiction principle. Well, it is common to hear of certain degrees of certainty, of more or less certainty, used in an extraordinary easy and natural way.

As Millan explains in (Millan, 2002, p. 116): "certainty derives from the Latin term 'certitude' which means 'strength'. In Latin, strength means 'stability', 'solidity', 'continuity', that is, just the opposite to hesitation or fluctuation. Whereas, the Latin verb 'certare' means 'to contest' (from here the term 'certamen' (contest) is derived), and 'certum' means what is decided or solved after some dispute. The term 'sure' is used as equivalent to 'certain' and etymologically means unconcern or careless, therefore, without fear. In this sense, to be sure is only a faculty for beings capable of knowledge and emotions. In the case of human beings, due to their sensory and intellectual capacities, there is a special fear called 'fear to err' … The absence of any fear to err is the state of mind called 'certainty'."

Therefore, a mind cannot be, at the same time, in two contradictory states as far as the truth of a judgement is concerned: its state has to be either certain or uncertain.

Something different is to deal with different types of certainty according to what human reasoning is able to do: deductive and inductive. And considering the latter and taking into account the strength of the laws which make up one of the premises of the syllogism: a natural law differs from a moral law. Classic philosophers distinguished among metaphysical or absolute certainty (for example, that of mathematics, or by means of sensible experience), physical, conditional or hypothetical (by means of laws known by experimental sciences), and moral (by means of ethical or social laws) for the above-mentioned reasons.

It is possible to reach certainty in those cases, and it is even valid to speak about a hierarchy in the strength of the truth of a judgement in mind depending on the type of certainty: firstly metaphysical, then physical, and lastly moral. This adjustment type in the strength of certainty is also called 'epistemological relevance' or 'quality of certainty'.

When in forensic disciplines speak about adjustment on certainty we do not use the distinction mentioned above, which derives from scholars, but a metonymy is present. In this case the term 'certainty' is used instead of 'uncertainty', the last capable of adjustment, to designate the state of mind on the truth of a judgement in any situation. What is determined, the state of mind under certainty, characterised by absence of fear to err, is subsumed in the concept of uncertainty, essentially characterised by the opposite. Again as a part of something, certainty, is taken for the whole, certainty plus uncertainty, therefore there is a synecdoche.

Aristotle defined the opposed concepts as “those which mean attributes which cannot inhere in the same subject at the same time” (Aristotle, Metaphysics). And he distinguished four opposition types: contradictory, privative, contrary, and relative.

The contradictory opposition appears when a concept is just the negation of the former. This opposition is the root of the others because they all suppose that something cannot be its opposite. This opposition is derived from the principle of non-contradiction both in the knowledge and in the being of real things. For example: animal / non-animal; white / non-white; etc.

The privative opposition is the negation of an owed characteristic to a subject. For example: true / false; good / bad; etc.

The contrary opposition appears when there are forms of the same type. Both forms imply a determined perfection and allow intermediate degrees. For example: cold / hot; etc.

The relative opposition appears when there are two positive concepts both mutually exclusive and interdependent. For example: father / son.

What type of opposition is there between certainty and uncertainty? Since intelligence tends to naturally know reality it tends to achieve the certainty state. Uncertainty is, in this context, a deprival of an owed characteristic to intelligence. Therefore, the most appropriate opposition in relation to the pair of concepts (certainty / uncertainty) of our interest is the so-called privative. As it has already been mentioned all opposition is contradictory, therefore, certainty cannot be confused with uncertainty.

Is privation adjustable? Adjustment appears in the opposition of contraries, above all in qualitative perfections, and in this kind of opposition extremes are perfections. It is not the same in privation in which there are extremes but while the former is a perfection, the latter is just the opposite. In the case of certainty, this is considered to be a perfection, and consequently it cannot admit adjustment. However, uncertainty is adjustable because it is correlated with a state of mind concomitant with the degree of firmness with which it assents a judgement as true, namely there is a simple inverse proportional relationship between the said degree of firmness and the degree of uncertainty.

The term ‘certainty’ has the same meanings in English and Spanish. It means the highest strength of the assent carried out by mind on the truth of a judgement or when making a decision. It is wrong to speak about adjustment on certainty in both languages.

9.4.2 Doubt, conjecture, and opinion

If we call the property of a known object able to produce the state of mind or intelligence we call certainty evidence, when the known object is not perceived by intelligence at the maximum degree of clarity existent for the evident things, i.e. when it is for example perceived as something only possible or probable, then our mind agrees with strength below the certainty degree. We are in states of greater or lesser uncertainty as doubt, formally considered as the suspension of judgement, conjecture, a state between doubt and the explicit expression of the judgement, or opinion, a judgement reserved due to the existence of the

fear of being wrong. Therefore, uncertainty as a state of mind is the subjective effect of a real or logic relation between the cognoscent individual and the known object.

9.4.3 Evidence and logic truth

The philosophical concept of evidence is one of those present in the stormy philosophical debate about the theory of knowledge for ages. Depending on the philosophical opinion viewpoint chosen, the concept has been differently defined and even confused with other terms.

Evidence is a synonym of 'clarity'. The etymological origin is Greek and its original meaning refers to the clarity of what is bright or absolutely clear.

Understanding truth as the correspondence of a thing to the intellect, it is essential to know whether such a correspondence has been made or not, or in a more thorough way, whether we are able to measure the conformity degree. When such conformity is absolutely clear, when the clarity of the judgement is full, we speak of evidence. If not, we speak of what is possible or probable.

Therefore, the logical truth, the one which is perceived by our intelligence when acting, does not consist of the clarity of a judgement, but we wouldn't be able to know whether that judgement is true or not without such a clarity (evidence). This is often called 'truth criterion' because it is the channel through which truth appears absolutely clear to intelligence.

Evidence perceived by senses is called immediate, but when we discuss with a deductive reasoning, the logical conclusion from the premises is perceived by our intelligence as evidence, in this case intellectual, and in a mediate way. The clarity in the intelligence confirming the correspondence between what is thought and what is real (or, where appropriate, what is just merely thought) in a proof has the same strength as that in a sensible experience.

Arana comments in (Arana, 2012, p. 33) that "Both rationalists and empiricists were trapped into their own trap to hunt truth because they were fascinated by the equation that equals truth to evidence, and evidence to certainty... Definitely, certainty is turned into a premise of evidence, and the last one into a requirement for the truth. The truth-evidence-certainty axis swings around its subjective end and so the man is turned into measure of all kinds of things that have to do with a rigorous knowledge." This description of the terms truth, evidence, and certainty has the typical semantic burden of immanentism.

The distinction between intrinsic and extrinsic evidence is classic (Vernaux, 1971, pp. 150-154). The former is characterised by the direct appearance of the truth of a judgement before the intelligence. The latter is different. As classic example of extrinsic evidence we can quote to give evidence. In this case what was testified cannot make what has happened evident. Testimony cannot show nor prove what has happened, but it can only guarantee that such an event has happened. The truthfulness of the testimony could be evident in that it can be shown by other means of proof (for example, by similarities observed among different and independent witnesses). A good example to illustrate how to reach evidence in a mediate manner is that related to a historical judgement. Nevertheless, it is important to note that extrinsic evidence needs the help of the will, opposite to intrinsic evidence. For those who are ignorant of some historical events, the rejection of the will to accept a determined historical

assertion cannot be catalogued as an irrational behaviour. However, this could be the case of those who did not accept the conclusion of a proof or the existence of something known by sensible experience.

Aristotle-Thomas general epistemology states that (Vernaux, 1971, pp. 124-129):

- Logic truth has no different degrees as its ultimate foundation is existence. Logic truth is indivisible as for a specific judgement there is no middle way between the adjustment of it or not to reality.
- Logic truth lies on the intelligence.
- Logic truth or error only exists when something is stated. Logic truth formally lies on the judgement of something being this or that (the union of concepts in judgement has an existential sense: the union is real). Thus, in judgement there is a reflection which perceives the conformity of the attribution to being.

A sort of existentialism as the one described above can understand anything that is real, doubtful, rejected or imagined in relation to existence. If we state that a chimera does not exist, this is a truth based on existence. Therefore, uncertainty is understood in this article from the scope of philosophy of being on which we are based on, and this is a very relevant point to understand it.

9.4.4 Science, belief and faith

It is worth mentioning the difference between science, belief and faith. This difference is well expressed by the verbs 'to know' and 'to believe'. Science comprises all those cases where the judgement is determined by the object. When it is not, we face an opinion or an act of faith.

The term 'belief' shows the described features on Section 9.2.1.

Belief-certainty or faith may take two forms (Vernaux, 1971, p. 139): (i) one that is not based in any intellectual reason beyond the fact that we wish things to be the way we consider them, as it happens in all conceptions of faith called fideists that are characterised by lacking any justification, showing practical contempt towards human reason and rejecting the power to achieve the truth at a metaphysical level; and (ii) the form based in objective reasons which do not force, i.e. reasons that are not enough to convey consent and therefore will is needed to achieve it. This way of understanding faith can be summarized saying that we would never believe if we did not want to, though we think that it is necessary to believe. In this context, the most solid and well-founded faith can be reduced to the status of opinion by those who do not share it, particularly those who only accept the scientific certainty as genuine. Additionally, on the other hand, it is worth mentioning that this faith we are talking about is not exclusively related to religion, as the exercise of faith by human beings in their ordinary life is a common occurrence.

It is convenient to underline that any act of faith is based, in the last analysis, on an act of knowledge. It makes no sense to think of an infinite chain of acts of faith because, in such a case, it would be a case of fideism. Faith based on reason is based, therefore, on the assumption that it is based on a true knowledge.

9.4.5 Elucidation of terms: certainty, opinion, probability (subjective), evidence and belief

The causes of certainty are, in short, evidence and the authority of a wisdom which deserves our confidence, the former could be called rational certainty (physical or metaphysical) and the latter certainty-belief or moral certainty. If there weren't any of those sources of certainty we would be before an opinion.

The concept of subjective probability is often used in the specialised scientific literature about probabilistic inference in forensic science. This concept of probability is particularly useful in forensic science because an exclusive classic use of the frequentist meaning would prevent us from assigning the probability of occurrence of investigated facts within the criminal jurisdiction. These kinds of facts are unique and unrepeatable, like the concurrent circumstances that happened meanwhile.

For that very reason, a language that Lindley describes as 'personalistic' is used in such an inference, as it has been pointed out earlier, and it is apparent, for example, by consciously selecting the term 'belief' instead of 'opinion' (Lindley, 2007, p. 12). 'Probable' – it is said - means "the degree of belief the fact finder entertains that a certain fact occurred" (Taroni et al., 2006, p. 1). Probability is equivalent to an adjustment of belief in the occurrence of a certain fact. The measurement of the uncertainty about the occurrence of a certain fact is expressed in terms of degree of belief.

Besides, the adjective 'personal' is used to underline the fact that each person assigns the probability of an event from what he/she 'a priori' knows or assumes, from which it could be very probable to support two different but well-founded probabilities of a same fact. The 'personal degree of belief' in the occurrence of an event presents itself as the description of the supported concept of probability: the subjective one.

The term 'belief' is polysemic, as mentioned before, and it could be ranked with an opinion according to one of its meanings. Any opinion, by definition, is uncertain. Otherwise, it wouldn't be an opinion, but a certainty or the sceptical doubt. Does it make sense to speak about degrees of opinion? We can speak about a greater or lesser uncertainty when giving an opinion, but we do not adjust the opinion because, simply, we state it. This is a particular type of judgement, and what is adjustable is the strength, assertiveness or the degree of confirmation subjectively assigned with respect to its truthfulness. Therefore, it is believed that speaking about degree of belief as equivalent to degree of opinion is inconsistent.

However, in common speech we accept to speak about degrees of belief as equivalent to degrees of uncertainty (it is the case of expressions which start with 'I believe that …'). In this case, to add the adjective 'personal' seems superfluous because uncertainty cannot exist by itself without a specific mind and after making a determined act of knowledge.

Subjectivity supported by experts in probabilistic inference in forensic science for the concept of probability is something, in our view, fully justified and very well-applied in the mathematical logic developed by them (Taroni et al., 2014; Biedermann, 2015; Lindley, 2007; Aitken and Taroni, 2004; Taroni et al., 2001). However, the language that they use related to the theory of knowledge might not bear the rigour needed, in our view again, because perhaps the above-mentioned ontological foundations and linguistic problems which make the forensic expert-Court communication and the effective application of scientific advances in forensic

statistics in the judicial context difficult have not been borne in mind, when all is said and done.

It seems to be reasonable to avoid a terminology such as 'personal degree of belief' to define what is 'probable', not even – in our view – with the wise intention of pointing out its subjective meaning, when if we understood correctly the statement 'degree of uncertainty' we would achieve the same aim. The same reasons that could justify a more generalised 'degree of belief', that is, those based on probabilistic data that the vast majority could share due to their frequentist nature (Taroni et al., 2006, pp. 20-21), would be applied with the same effectiveness as if we spoke about 'degree of uncertainty'.

In this sense, general epistemology or theory of knowledge may offer an important contribution to probabilistic inference in forensic science. Terminological precision reached after a lot of years of philosophical discussions may be a good ally provided that those terms are defined within a reliable epistemological school of thought as the Aristotle-Thomas realism in order to guarantee internal coherence.

The time has come to criticise – we believe it to be constructive – regarding the use given to the expression 'degree of belief' in the quoted book written by Lindley. He argues that a term is needed to describe what we feel about an event and says that the phrase usually used is 'degree of belief'.

Then, he adds that we will have the highest belief when we think the event is true, and the least when false. And he points out the usefulness of the term 'belief' because it emphasizes the fact that the uncertainty we are talking about is a relationship between us, on the one hand, and an event, on the other hand.

He also says that belief does not reside entirely with us because it refers to the world external to us, that it is not a property of that world – he states – because the degree of belief of one person may reasonably be different from another. The term 'belief' expresses, according to Lindley, a relationship between us and the world, in particular between us and an event in that world. He concludes this part of his dissertation by saying that probability is used to measure the strength of our belief in the truth of an event.

Lindley says that probability does not exist in the external world on its own, supporting de Finetti's statements[28]. Nevertheless, we think that the terminology used by Lindley to justify the concept of subjective probability is unfortunate.

[28] A. Biedermann (Biedermann, 2015, p. 145) quotes the following words written by de Finetti: "…'Probability does not exist' by which I mean that probability does not 'exist' on its own, independently of the evaluations we make of it mentally or instinctively", and just in the end says: "The question 'what is the probability (…)' is not only defective because of the 'the', but also because of the 'is'". On page 144 A. Biedermann states that: "It represents a judgement made by someone based on his/her assessment of the available evidence". If we stick to the linguistic concept of uncertainty, we know it refers to a state of mind regarding the truth of a judgement. It deals with the intellectual operation performed by an individual capable of such a task and also with the results of the operation itself. Even more, it deals with the unsatisfactory result of such an operation as the intellect aspires to rest in the possession of truth when it recognizes the reality. Uncertainty is a being whose existence is only in mind, though it comes from the direct knowledge that human beings have of reality in the last analysis. The measurement of such an adjustable loss is conceived as a subjective judgement by those who defend the so-called subjective notion of probability.

Considering the three possible meanings of 'belief' (general judgement, judgement with reservations or opinion, and certainty achievable by an act of faith) the meaning used by Lindley is different because it is referred to as conviction or confidence. Such a personal conviction (adjustable) is directly related to all kinds of possible uncertain and certain (true or false) states, also with the firmness of assenting to a judgement. There is a direct relationship with respect to the degree of firmness, and the adjustment is exactly the same. There is an inverse proportional relationship with respect to the degree of uncertainty.

On the other hand, the chosen sentence (degree of belief) may cause confusion when it is thought to be adjustable. Judgements are not adjustable but our conviction of their adaptation or not to reality (this is truth as a property of a judgement) and such a degree of firmness in the assertion of a judgement (or, in other words, the degree of personal conviction, subjective, of the truth of a judgement) is described by the words: certainty and uncertainty, therefore, why do we need to find out new ways to say the same thing? Probability is the tool that we usually use to measure such a degree of conviction.

An exception may be made to the above-mentioned reasoning if the meaning of 'belief' is certainty-belief. In this case it is reasonable to think of certain adjustment regarding confidence inspired by authority in those who are concerned, but such an adjustment is similar to the epistemological degree of certainties according to the law (metaphysical, physical, or moral) supporting them. It would be an adjustment of a different kind than that claimed by the expression 'degree of belief'.

In any case, with certainty-belief the will comes into play necessarily. This is the essential difference between knowledge and belief, between the verbs 'to know' and 'to believe', as mentioned above. The achieved certainty-belief does not come from objective or intellectual evidence, mediate or immediate, perceived by intelligence in the known object, as it happens in science. The assent in believing comes always from the will.

The problem detected using the term 'belief' and its derivative 'degree of belief' is the semantic bias given by the scientific literature related to the theory of probability, basically from the appearance of the Bayesian inference approach.

Belief has been identified with subjective judgements and the degrees of belief with a specific way of understanding probability, namely with the so-called subjective probability. Statements and propositions are considered to be objects of belief – i.e. beliefs – and it is said that a way of measuring uncertainty may be by means of belief functions (noted by Cred), with arguments p (proposition) and s (subject), whose values are included in the closed interval [0, 1] which represents the 'degrees of belief' that a subject s has in the proposition p: Cred (s, p) $\rightarrow$ [0, 1], being 0 the null belief, 1 the full belief and any other value the partial belief (Velarde, 2005, p. 28).

The Bayesian inference approach enables the combination of probabilities based on data (also called objective or frequentist probabilities) with subjective probabilities (assigned by considering experience, knowledge and information) (Aitken and Taroni, 2004, pp. 21-22). What is relevant here is the gradual consolidation in the scientific literature about evidence evaluation by forensic experts of the similarity between the concept of belief and the subjective judgements, on the one hand, and the measurement of uncertainty of such subjective judgements (statements and propositions) and the measurement of belief of such

beliefs (the same statements and propositions), on the other hand. It is an ambiguous way of understanding the term 'belief' that may cause, in our view, misunderstandings to distinguish well between the verbs 'to know' and 'to believe'.

It is suggested changing the sentence 'personal degree of belief' for 'personal degree of credibility'. Though the former is the common expression in the specialized literature, according to the etymological meaning of 'credibility' it comes from the Latin word 'credibilis' meaning quality of credible. Likewise, 'probability' comes from the Latin word 'probabilitas' and means quality of probable, something that may happen. A concept is defined as a quality of another when it refers to a determined property. In this case we have referred the properties (i.e. qualities of 'probability' and 'credibility') to a judgement, being so this judgement both probable and credible. What is probable and credible is the judgement. But how much probable / credible is it? This is determined by means of a degree of probability / credibility. And the term 'personal' has to do with what a person 'a priori' assumes and knows before asserting a judgement.

'Belief' as judgement is not a quality. Beliefs are stated, without further ado. However, when we speak of 'credibility' we refer to a quality, a property that a belief has for us: namely, to what extent does that belief deserve credit for us? So, as measure of belief, credibility is adjustable and then it is possible to establish degrees of 'credibility'.

What is the scope of the 'degrees of belief' used in forensic science together with the concept of subjective probability? What is the nature of such subjectivity? Taroni et al. responds – using De Finetti's words – that belief is made up from knowledge, experience and information (Taroni et al., 2001, pp. 145-150). If subjectivity is related to those three factors it is believed that the proper word is opinion (assent with reservations or, in other words, with uncertainty), but not belief, and the firmness of the assent is correctly expressed in terms of degrees of uncertainty or credibility.

Therefore, it seems unfortunate to use the expression 'degree of belief' to describe the nature of the subjective probability. Though this has been made with the good intention of pointing out that probability is a state of mind, and not a state of nature following a famous Savage's expression, it is considered that the term probability does not need to be qualified to explain its true nature from an Aristotle-Thomas realistic epistemology. The answer is a suitable understanding of the nature of an act of knowledge: it has, basically, epistemological nature.

While probabilities that could be assigned from actual events are all of them conditionals, there are no human acts of knowledge without previous knowledge, information, and experiences. And within them there are some from immediate and mediate origin (most of the latter are those beliefs based on the authority where knowledge, information or experience comes from). Therefore if the expression 'degree of credibility' is preferred than 'degree of uncertainty' is not because the subjective concept of probability is underlined but because the actual knowledge of a happened event is a mixture of knowledge coming from the direct action of intelligence on the matter at hand and that based on the authority which knowledge comes from.

9.4.6 Cause, effect, principle, condition, and occasion (Metaphysics, book X; Vernaux, 1971; Alvira, 2001; Corazon, 2002; Millan, 2002; Llano, 2007)

The notions of cause and effect are innate in human beings in their ordinary experiences, basically with respect to what is called efficient causality:

- External[29]: we distinguish temporal series of events that happened before us and causality very well. We do not consider the fact of an event following another event, as the summer follows the spring, to be a cause-effect relationship. It is well-known that such a sequence of events is produced by the movement of heavenly bodies.
- Internal: we are all aware of the power of our will over our internal potencies or faculties (memory, imagination, discursive capacity, and so on). It is classically known as ab intra ad intra experience.
- Internal-external: we are also aware of our action upon things or people around us, or how they influence on us. They are classically and respectively known as ab intra ad extra and ab extra ad intra experiences.

Hume reduced the causal influence to a mere chronological sequence of phenomena. If we define science as accurate knowledge through causes and causality is denied, it will be unavoidable to fall into some way of scepticism. Hume did not deny causality, properly, but the possibility to know causes from observations of phenomena. Causal propositions are not verifiable for him. He supported that men have the belief of their existence because they observe chronologically successive phenomena, and think that they are interdependent.

However, the temporal continuity observable in many causal phenomena is not a necessary condition for every cause. For example, there are different hierarchical levels of responsibility in professional life, but it makes no sense to support that the attaining of each level is cause of that of the following in the hierarchy order but, in any case, a condition. These concepts cannot be confused: cause, principle, condition, and occasion.

According to Millan, "the most proper and rigorous meaning of the term cause is that established by consecutive determinations of the notion principle" (2002, p. 75). If we call that from which something is coming principle, the way of coming from is not determined in its concept. The concepts of cause and principle are different due to such a determination on how they come from.

However, Millan distinguishes between real and logic principles before going into explaining the concept of cause. The premises of a syllogism are logic principles of the conclusion. The conclusion is inferred from the premises but they cannot be considered to be their real principles.

[29] A. Millan (Millan, 2002) states that: "the use of the principle of causality, when exterior beings are considered, is not an ontological issue but methodological, which is solved, when there is enough experience, by applying the laws of induction, and taking into account all kinds of necessary conditions in order not to proceed without foundation. And in those cases wherein there were no basis to make an induction rightly, the principle of causality would continue to be valid: 'to be an effect' is not the same as 'to be known as an effect', likewise 'to be' is not the same as 'to be known'".

Millan also distinguishes between positive and negative real principles (2002, pp. 76-77). The former can bring or not something to be, or to be or not the foundation of the becoming of that which they are principle. For example, the steps of a preconceived plan to commit a crime have a chronological order, but the first step does not bring the following steps to be, nor it is the foundation of its becoming even being the principle of all of them. The fact that a real and positive principle might become a cause, it has to bring something to be or to be the foundation of its becoming, being such a thing really different and dependent on it. So, the above-mentioned plan is really a cause because without it the criminal activity developed in practice would not have been produced (note here the dependency on the becoming), making up a real positive principle in this case. This kind of cause is called final cause. The criminal activity consisting of shooting victims is also an example of a real and positive principle because it brings something different to be (the quality of innocent victim of that who was shot). The victim would not have been such if the criminal had not shot him / her. This type of cause is called efficient cause.

A negative real principle is a lack or loss. We cannot speak about a concrete death unless the subject at hand is not alive beforehand. The lack of life makes up a negative real principle, being something extra-mental necessary for the death to be real. However, the lack of life is compatible with a lot of possible explanations to cause the death therefore it is not the cause of it.

Therefore, cause is every real and positive principle from which something is coming and dependent on being or becoming. There must be a real distinction between the cause and the effect. There is a clear priority of the cause over the effect, and this priority is related to the existence of a natural order: that perfection that the cause produces in the effect has to be previously in the cause. This priority will be, in many cases, a temporal priority, but such a class of priority is not essential for the concept of cause.

The following properties which characterise the cause-effect notion[30] can be pointed out (Alvira, 2001, pp. 207-208):

- Effective dependence on being or becoming: a cause will be such if the effect is impossible to become or to be;
- Real distinction between cause and effect: a real dependency between two things implies to distinguish between the two of them necessarily;
- Priority of cause over effect: every cause is previous to its effect according to a natural order. Perfection conferred or produced in the effect by its cause has to be found, somehow, previously, in it. This priority does not imply to always understand it from a merely temporal viewpoint.

It is only possible to understand well the principle of causality of the philosophy of being starting from its concept of being: the notion of being is not a genus; it is not possible to add any difference not previously included on. This property of the notion of being is called

[30] D. Lindley (Lindley, 2007, pp. 57-58) underlines the necessity to distinguish between 'association' and 'causation' when we are interested in developing the concept of our probability of an uncertain event E, given that we know or suppose another event F to be true, that is, p(E | F), describing the former as 'seeing', and the latter as 'doing'. Causation has been exclusively understood as causal efficiency by some philosophers. Leibniz thought that action is the essential character of substances, and not something accidental. Of course, Lindley's argument on probabilities in that paragraph is very relevant from the logical viewpoint (see also footnote number 29).

extension. From this aspect, the notion of being has the maximum possible extension, covering all kinds of realities. But, besides, its notional content is also the maximum because it means all kinds of realities with their singular characteristics. This property is called comprehension.

To announce a singular reality, it is not enough to say that it is a being, but it is necessary to determine its essence, its way of being. Nonetheless, it is possible to say that all realities are beings, and this statement is based on the fact that all of them have being.

The notion of being as genus has been used by philosophers who understand the metaphysic concepts as logic ones. According to these philosophers what intelligence knows firstly are singular beings. Afterwards, they know essences by abstraction. And lastly, above all of them, they know the most abstract concept, being. So, being is understood as the most generalised idea of being. The consequence of this way of understanding being is to hold that metaphysics does not have to do with experience and reality.

The principle of causality has been formulated in many ways, but the deepest one is that stated by St. Thomas Aquinas in his Summa theologiae: “everything which is advisable to something and does not belong to its essence, belongs to it by some cause.” Knowledge that a human being acquires during his life has a cause outside his own nature: parents, teachers, books, and so on. Applying this formulation to being, all beings agree on being, not belonging to their essences, and, consequently, it demands a really different cause. Essence is by its own nature principle of diversity. And, in consequence of understanding the principle of causality in this way, it follows that “nothing can be cause of itself,” because otherwise the principle of non-contradiction will be infringed.

M. Artigas (1992) states that: “The principle of causality expresses a general condition which is accomplished in reality, and therefore, is also a condition of possibility of our knowledge: it refers to the necessity that all that happens is the result of causes. The particular causes, which can be of different types, are not specified; if they are necessary or free, determinist or not, neither; it is only asserted that the natural results must be the result of causes. It is a modest assertion regarding details, totally general in its scope, and completely certain. It cannot be proved by the procedures of experimental science, but this fact does not prevent it from its validity; on the contrary, the validity of any knowledge, included that of science is based on the principle of causality. It is not merely speculative either, because it is the result of the analysis of what experience shows us, though transcends what is given in the experience because it leads to the rational explanation of reality.”

It suits us to distinguish the concept of cause from the remaining mentioned concepts in the paragraph: principle, condition, and occasion.

Principle is that from which something comes anyhow. Therefore, every cause is principle but not vice versa. Principle implies beginning or order. Cause is so, a type of principle. Privation can be considered a peculiar type of principle, that is, the absence of perfection as starting point to acquire that perfection.

The condition is the needed requirement to perform causality. Those requirements make possible or prevent the performance of a cause, but the condition as such has no causality.

An occasion is such a situation which is beneficial to the performance of the cause. However, it is not a necessary condition for the cause to act.

9.4.7 Determinism, indeterminacy, and chance

According to the philosophy of being, the objective chance (supported by somebody as existing out of our mind) is empty of being, and the subjective one (that which we think of) is empty of knowledge (Arana, 2012, pp. 152-153). Physicists are entitled to speak in terms of chance and necessity, but metaphysicians need to find a less vain partner of necessity to explain reality (understanding it as what is out of mind), and that partner is called contingency, that is, the existence of non-necessity in reality. The term contingency has to be understood as it is expressed by the philosophy of being: what is contingent is the being without the act of being by itself, but the act of being received from another. There is only one necessary being from this philosophical school, that who is by himself, that is, God.

Determinists have got to domesticate chance, according to Hacking's words, allow themselves to speak with perfectly ordinary voice on laws of chance – which is rather a remarkable paradox – it is equally a part of reality subjected to their deterministic ideology. That chance is a comfortably accepted reality by determinists because it is neutral, that is, allows them to explain the complexity of reality without having to admit the final cause.

Arana points out intelligently that "chance is not 'the other' of necessity, but rather 'the same' but negatively expressed." Contingency – Arana says – is "an idea that cannot be reduced to necessity turned upside down, but it points out to something truly different and therefore much more laborious to be expressed in terms of numbers and laws" (Arana, 2012, p. 150).

"Chance, in its relative sense, is a concept without an own substantive content, designing what is left without determination by certain epistemic domain (discipline or theory)" (Arana, 2012, p. 60). In other words, "Existence in itself demands a perfect definition of all the features that shapes it" (Arana, 2012, p. 59), therefore, "The 'ontological indeterminacy' cannot be acceptable…" because it infringes the principle of non-contradiction. Those who defend the real existence of ontological indeterminacy (objective chance) are scandalized by those who explain such realities by means of contingency.

9.5 Conclusions

Scientificism despised spontaneous knowledge (Artigas, 1992; Corazon, 2002). Pre-scientific experience in ordinary life is necessary to live and behave. The human being is a social, cultural, and historical being. This means that tradition and teaching are true sources of knowledge needed for living and behaving and human faith is absolutely indispensable as well (Corazon, 2002). All these sources of knowledge and a precise use of gnoseological terms are thought to be needed for expressing the real value of knowledge about facts reached in penal proceedings.

As a result of the previous discussion it is apparent that Bayesian inference may be the only reliable tool to make probabilistic inference and answers most difficult questions relating to reporting in forensic science. However, the judicial language used in courts to prove statements on facts needs a minimum of philosophical realism as some European legal scholars are claiming (Gascon, 2010). Therefore, concepts such as truth, certainty, doubt, evidence, belief, and others, need to be clarified from such a basic philosophical perspective, or at least, need to be defined according to the philosophical view in which they are based on, and it has not been carried out in forensic guides so far.

APPENDIX II

Royall says the following: "Suppose A implies that X has the probability distribution $p_A(\cdot)$, while B implies $p_B(\cdot)$. If B is true then when we observe X it is unlikely that we will find strong evidence favoring the false hypothesis A. Specifically, for any given constant k > 0, $\Pr(p_A(X)/p_B(X) \geq k) \leq 1/k$.

This is because, if S is the set of values of x that produce a likelihood ratio (in favour of A versus B) of at least k, then when B is correct: $\Pr(S) = \sum_S p_B(X) \leq \sum_S p_B(X)/k \leq 1/k$.

The first inequality is obtained because, for every x in S, $p_B(x) \leq p_A(x) / k$, and the second because the sum $\Sigma_S p_A(x)$ is the probability of S when A is correct, which cannot exceed one.

A similar argument can be used to prove a much stronger result:

Let X_1, X_2,… be independent and identically distributed (i.i.d.) random variables, and p_A and p_B conditional probability distributions to the events A and B respectively. What is necessary to be proved is the following: when B is true, the likelihood ratio in favor of A converges to 0 with probability 1.

Proof:

From the inequality (1.3) on page 7 of the Royall's book:

$$\Pr\left[\prod_1^n \frac{p_A(X_i)}{p_B(X_i)} \geq k \ \textit{ for some } n = 1,2,\ldots\right] \leq 1/k \Rightarrow$$

As Pr(event) = 1 – Pr(opposite event):

$$\Pr\left[\prod_1^n \frac{p_A(X_i)}{p_B(X_i)} \geq k \ \textit{ for some } n = 1,2,\ldots\right] = 1 - \Pr\left[\prod_1^n \frac{p_A(X_i)}{p_B(X_i)} < k \ \textit{ for some } n = 1,2,\ldots\right] \leq 1/k$$

$$\Rightarrow 1 - 1/k \leq \Pr\left[\prod_1^n \frac{p_A(X_i)}{p_B(X_i)} < k \ \textit{ for some } n = 1,2,\ldots\right] \Rightarrow$$

Taking logarithms and multiplying by 1/n:

$$\Rightarrow 1 - 1/k \leq \Pr\left(\frac{1}{n}\log\left[\prod_1^n \frac{p_A(X_i)}{p_B(X_i)} < \frac{1}{n}\log(k) \textit{ for some } n = 1,2,\ldots\right]\right) \Rightarrow$$

Logarithms turn products into additions:

$$\Rightarrow 1 - 1/k \leq \Pr\left[\frac{1}{n}\sum_1^n \log\frac{p_A(X_i)}{p_B(X_i)} < \frac{1}{n}\log(k) \textit{ for some } n = 1,2,\ldots\right] \qquad (*1)$$

Applying the law of large numbers:

(the mean of i.i.d. random variables converges on the mean of them with probability 1)

$\frac{1}{n}\sum_{1}^{n}\log\frac{p_A(X_i)}{p_B(X_i)}$ converges with probability 1 on the mean of $\log\frac{p_A(X_i)}{p_B(X_i)} = E\left[\log\frac{p_A(X_i)}{p_B(X_i)}\right]$ by the Jensen's inequality $E\left[\log\frac{p_A(X_i)}{p_B(X_i)}\right] \geq \log\left[E\left[\frac{p_A(X_i)}{p_B(X_i)}\right]\right]$.

Taking limits when $n \to \infty$, the right term $\frac{1}{n}\log(k)$ of (*1) tends to zero. Therefore, the left term $\frac{1}{n}\sum_{1}^{n}\log\frac{p_A(X_i)}{p_B(X_i)}$ converges to the constant $E\left[\log\frac{p_A(X_i)}{p_B(X_i)}\right]$ and $\log\left[E\left[\frac{p_A(X_i)}{p_B(X_i)}\right]\right]$ is a negative constant.

Therefore, $1-1/k \leq \Pr\left[\log\left[E\left[\frac{p_A(X_i)}{p_B(X_i)}\right]\right] < 0\right]$

Being a logarithm involves that the quotient between probabilities is lesser than one on average. This implies that the strengh of evidence supporting B over A will be higher than any previously fixed value k.

As this has been done through the limit process, for that value k and any other randomly small number $\varepsilon > 0$, it is possible to select a large enough value of n that there is a probability higher than $1 - \varepsilon$ that for this n to find k times evidence supporting B over A.

As Royall says, this ensures that given any number k, for example 100.000, we will always find a number n that the probability of obtaining 100.000 events is in favour of the true hypothesis with probability almost 1 (convergence to 1 in the limit).

APPENDIX III

Bayes' rule (Taroni et al., 2006, pp. 6-7):

"A relevant proposition B, with respect to a proposition A, is a proposition such that if it were true, then such information would change our degree of belief in A... For any propositions A and B, the degree of belief that is true, given that one assumes that B is true, is equal to the degree of belief that A and B are both true, given background information I, divided by the degree of belief that B is true, given background information I, provided that Pr (B | I) > 0:

Pr (A | B, I) = Pr (A, B | I) / Pr (B | I)." This albebraic version of the formula first proved by Reverend Thomas Bayes is called *Bayes' theorem*. And the authors underline: "The importance of Bayes' Theorem is due to the fact that is a rule for *updating degrees of belief on receiving new evidence*."

Jeffrey's rule (Taroni et al., 2006, pp. 19-20)

Let's consider all the possible scenarios that can be derived from the combination of two logically compatible propositions A and B at time t_0. Then, at time t_1, proposition B is known to be true. The authors show that the use of Bayes' theorem for probability updating is equivalent to the redistribution of probabilities among possible scenarios in a symmetric way; given that the state of information has changed on learning only that B is true, and nothing else, there is no reason to make a change biased for and against certain particular scenarios.

When reasoning under uncertainty, the probabilities of the four possible scenarios at time t_1 are greater than zero. How can be distributed the probabilities of scenarios in such a way that they add up to one to accomplish the additive rule of the theory of probability? A reasonable answer is that the probabilities are redistributed such that the ratio between the new and the old probabilities of the scenarios is the same as the ratio between the new and the old probability of B because it was the only change in our personal state of information:

Pr_1 (A, B | I) / Pr_0 (A, B | I) = Pr_1 (B | I) / Pr_0 (B | I); from this formula, this is the rule for calculating the new probabilities: Pr_1 (A, B | I) = Pr_0 (A, B | I) × (Pr_1 (B | I) / Pr_0 (B | I)). The last term is a constant factor, and the probabilities have been redistributed among the possible scenarios in a symmetric way: given that the state of information has changed on learning only the new probability Pr_1 (B | I), and nothing else, there is no reason to make a change biased for and against certain particular scenarios.

The authors say that this formula is "a straightforward generalization of Bayes' theorem, known in the philosophical literature under the name of Jeffrey's rule," because Jeffrey was who first argued that it was a reasonable general updating rule (Jeffrey, 1983).

APPENDIX IV Voice comparison Outline of the report	Links between the example statement and the standard (numbers refer to the relevant section in the standard)
Evaluative Statement	Evaluative reports should be identified as such (Section 2.2)
Information Contextual data known from the Intelligence Service: 1. A man made a phone call to a private call centre, in a terrorist organization's name, from a mobile phone warning that a van bomb had been planted with a large amount of explosives at a parking of a Terminal of an international Airport, detailing place, model, color, and registration plate which would go off one hour later approximately. 2. The explosion of the van bomb took place as planned by terrorists. The van was stolen by terrorists some days before in other country. 3. The terrorist organization claimed the attack few days later in a local press. 4. Two people were arrested in a Police road check after six days from the attack. In the napsack of one of them two guns stolen in the other country were found, and personal documentation of him contained a handwritten note with two of the three phone numbers where warnings were made by terrorists. 5. The terrorist command was consisted of three people.	Relevant case information as understood by the forensic scientist is disclosed as part as the requirement of transparency (Section 3.12 and Guide Note 1).
Items Received From the Investigating Judge: ▪ EMTEC C-60 cassette tape related to the judicial statement made by the main suspect (suspect1). ▪ CD with two bomb alert recordings related to the Terminal of the Airport bomb attack. ▪ 3 CDs with speech recordings obtained at prison centres from the three members of the terrorist command in wma format, and official letters signed by Judges authorizing penitentiary authorities to tap those conversations.	Indication of the items received as part as the requirement of transparency (Section 3.12 and Guidance Note 1).

To avoid possible damage on the original tape due to rewinding and stops, a Windows PCM WAV audio format digital copy is made. While performing this task we observe that the tape speed fits the standard of 4.7 cm/sec. Likewise, the WMA audio format of the evidence number 3 has been converted into a Windows PCM WAV format.

All the converted WAV audio files are sampled at 8 KHz, 16 bits, mono. The accuracy of this format ensures the integrity of the information available, being also compatible with the software (Automatic Speaker Recognition system) used in this report.

The audio files are edited to separate the unquestioned voices of the defendants and the remaining questioned voices, measuring the audio quality according to the criteria of evidence admissibility as established in the Technical Guide in force in this Department, which are summarized as follows:

Questioned voice and test files:

A. Audio duration higher than 10 seconds after the edition process to separate voice from background noise and regions of silence.
B. Signal to noise ratio higher than 10 dB.
C. Intelligibility of all sounds with respect to the background noise.
D. Non-existence of a significant distortion of the audio signal due to saturation or other causes.

Files used to train speaker models:

A. Audio duration higher than 40 seconds after the edition process to separate voice from background noise and regions of silence.

The remaining paragraphs (B, C and D) are the same as above.

The next table shows the unquestioned and questioned files after being edited under the earlier parameters:

Item	File	Duration (edited file in sec.)	S/N ratio (dB)	Observations	Brief nomenclature
001	Tape_Side_A.wav	> 300	7.3	Additive and convolutional noises, bad perceptive quality.	Suspect1_1
001	Tape_Side_B.wav	< 60	7.3	Additive and convolutional noises, bad perceptive quality.	Suspect1_2
002	Van bomb threat to call centre1_32sec.wav	32.7	25.5	Slight nasalization.	Alert1
003	Suspect1_prison_120seconds_microphone	125.49	31.5	Spontaneous voice.	Suspect1_3
003	Suspect2_prison_60seconds_landlinephone	75.52	36.2	Spontaneous voice.	Suspect2
003	Suspect3_prison_120seconds_landlinephone	121.31	24	Distortion due to channel noise, bad perceptive quality.	Suspect3
004	Van bomb threat to call centre2_17sec	17.2	5.2	Additive noise and distortion by the effect of the channel.	Alert2

According to the Technical Guide the following files can be used to train speaker models: • Suspect1_3 • Suspect2 And as test file: • Alert1.	
Issue Comparison between the three unquestioned suspects voices and the Terminal bomb alert.	The issues are here at source level because it is not risk of misleading if it is assumed that the issue of source (recorded speech in a call centre) is directly related to an activity (to speak by phone at the same time). See Guidance Note 2.
Nature of Examination It has been experimentally tested that the speaker's identity is strongly correlated with his/her physiological and behavioural characteristics (linguistic habits, intonation of sentences, etc.). Size variation of the vocal tract cavities gives rise to differences between their resonances. Size and elasticity variation of vocal cords derives in changes of the mean of the fundamental frequency of voiced sounds. Soft palate and size of the nasal cavity variations produce nasal sound spectral differences. Atypical anatomical variation (arrangement of teeth and dental health, palate structure, etc.) may give rise to atypical sounds or an abnormal nasality. This way each speaker will introduce certain peculiarities in the speech signal allowing listeners familiarized with his voice to identify him. Factors affecting intra-speaker variability may be divided into two big groups:	

▪ Voice signal variability: it is derived either from factors directly attributable to the speaker, voluntary or involuntary (the so-called intrinsic factors: age, emotional state, physical state, stress, articulation speed or speech style (read, whispered, conversational, etc.), or from external circumstances to the speaker (the so-called extrinsic factors: speech signal capture and transmission devices, bandwidth, channel distortion, reverberation, additive noise, etc.). ▪ Variability along the time. Essential features of the software applied to this forensic casework: It is a speaker recognition system that uses MFCC parameters, GMMs, channel normalization techniques (CMN, RASTA and FEATURE WARPING) and likelihood normalization techniques as well as in the recognition process. The system is adapted to the Bayesian evaluation of evidence approach calculating robust LRs. A main novelty with respect to earlier versions of the software is the D-norm normalization, which is based on the KL distance from the speaker statistical model to the UBM model. The calculus of the KL distance is approached by the Euclidean distance among mixtures. This software also includes an optional likelihood normalization related to the scores obtained comparing test files with the so-called test models. The system validation has been carried out by means of the participation of the research group in the NIST Evaluations (Department of Commerce of the United States of America) held from 2001 to 2008. The results obtained by the research group in the above mentioned Evaluations, were accessible to any research group in the world, and the results that were published in the proceedings of the main international conferences on speaker recognition have shown that the technological kernel of the software used is one of the best speaker recognition systems as classifier, as well as unique in its kind in the market, so far, allowing experts to calculate LRs as strategy to evaluate evidence in Court.	
Examination and results Firstly, statistical speech models are made from the voices of suspects using their unquestioned voices recorded on the CD-Rs labelled as Item number 003 from taps carried out in different prisons. Some experiments are carried out to check the suitability of such recordings to perform the voice comparison.	

They consist of a series of cross-matching tests among unquestioned voices, adjusting the reference population one by one and using the remaining as test files.

Next, the voice comparison is made in the light of two alternative propositions: that of the Prosecution relating the questioned voice recorded in the tapping process to the unquestioned voice from the suspect, and that of the defence in the opposite sense.

Technical characteristics of the tests carried out are described below:

Parameterization characteristics:

- Channel normalization: CMN + RASTA + FEATURE WARPING
- Coefficients: MFCC and delta
- Number of filters: 20
- Window type: HAMMING
- Window length: 20 milliseconds
- Overlapping (%): 50
- Amplitude: 1
- Pre-emphasis: No
- Sampling frequency: 8000 Hz
- Case-impostors for Z-normalization

Modelling characteristics:

Suspect model:

- Mixtures: 1024 (male voice-UBM)

Model characteristics:

- Training time: 60 or 120 seconds
- Session type: Mono
- Speech style: conversational

Test strategy:

The following comparisons will be made:

CT01:

- Suspect model: Suspect1_3 (Microphone channel).
- Reference population: Microphone-Spanish-Men, 35 best-competitor speaker selected, edited frames of 120 seconds. This population database comes from recordings from a big database belonging to the laboratory with recordings of 120s, obtained using digital wire-tapping.
  Public Database reference: BDRA public file (Order INT/1202/2011, 4th of May, Spanish Ministry of Interior, BOE nº 114, 13th of May, 2011)
- Test file (edited speech): Alert1 (GSM channel).

CT02:

- Suspect model: Suspect2 (Landline channel).
- Reference population: Landline phone-Spanish-Men, 35 best- competitor speaker selected, edited frames of 60 seconds. This population database comes from recordings from a big database belonging to the laboratory with recordings of 60s, obtained using landline wire-tapping.
  Public Database reference: BDRA public file (Order INT/1202/2011, 4th of May, Spanish Ministry of Interior, BOE nº 114, 13th of May, 2011)
- Test file (edited speech): Alert1 (GSM channel).

Due to the slight nasalization perceived in the questioned file, a series of audio test files with speech from people different to defendants, the so-called case-impostors, experimentally recorded simulating similar condition to the case at hand, i.e. the existence of nasalization and using a GSM channel, have been included.

Likewise other files different from the above-mentioned but recorded in similar conditions have been also used to apply the Z-normalization, a technique needed to strengthen the scores.

Checking of channel adjustment in the selected reference models:

As a consequence of the experiments carried out and according to the empirical conclusions achieved so far, the channel adjustment between the so-called "suspect model" and "reference models" is one of the most influential factors affecting the adjustment of all the scores in the experiment, in such a way that an adjusted reference population allows the expert to control it and validate the selected models.

The adjustment checking consists of observing that each test file (those from people different to the suspect but recorded on the same channel type) obtains a score between -2 and +2 within a normalised scale. The following table shows the achieved scores in the above-mentioned experiment.

	Scores	
	CT01	CT02
Impostor 1	-1,033674376	-0,55465494
Impostor 2	-2,471026857	-0,16844269
Impostor 3	-0,144728539	1,31029527
Impostor 4	-0,717954227	-0,33435225
Impostor 5	-1,398506885	-1,1415008
Impostor 6	-1,114991684	-0,14569758
Impostor 7	-0,417311084	-0,28217063
Impostor 8	-0,770303354	-1,8281457
Impostor 9	-0,842499005	-0,2479815
Impostor 10	-0,942573979	0,408688305

As it can be observed, almost all the scores are in the above-mentioned range and therefore, the reference population is considered to be reasonably channel-adjusted.

Channel adjustment between test files and models:

For this channel adjustment we use a series of files from people different to the suspect – the so-called case-impostors – recorded on channel conditions similar to those of the questioned files. By means of a t-student

test the system checks whether the score distribution obtained from the earlier files is adjusted to the expected distribution.

A further normaliser technique is used to strengthen the scores against the channel influence and, in this case, also specifically against the nasalization effect found in the questioned file. A Z-normalization is applied. The table below shows the results obtained.

	Scores			
	CT01		CT02	
Impostors	Before adjustment	After adjustment	Before adjustment	After adjustment
Impostor 1	-0,082	0,254	1,199	0,861
Impostor 2	-0,134	0,202	1,321	0,983
Impostor 3	-2,609	-2,273	0,090	-0,248
Impostor 4	-0,512	-0,175	1,753	1,415
Impostor 5	-0,431	-0,095	0,012	-0,326
Impostor 6	-2,608	-2,272	-0,644	-0,982
Impostor 7	0,229	0,636	1,269	0,931
Impostor 8	-1,331	-0,995	0,540	0,202
Impostor 9	-0,398	-0,062	-0,463	-0,800
Impostor 10	1,360	1,696	0,510	0,172
Impostor 11	0,210	0,547	0,880	0,542
Impostor 12	-0,041	0,295	-0,238	-0,576
Impostor 13	0,304	0,640	-0,719	-1,057
Impostor 14	0,399	0,735	0,270	-0,068
Impostor 15	0,530	0,867	-0,710	-1,048
Mean	-0,336	0,000	0,338	0,000
Standard deviation	1,096		0,814	

These data support that the obtained scores are reasonably adjusted to those of the expected distribution [~N(0,1)].

Results of the voice comparison:

CT01: Model: Suspect1_3

FILE	LR
Alert	213,046
IMP 01	0,027
IMP 02	0,025
IMP 03	0,494
IMP 04	0,077
IMP 05	0,053
IMP 06	0,019
IMP 07	0,018
IMP 08 (MIRRA 01)	0,128
IMP 09 (MIRRA 02)	0,506
IMP 10 (MIRRA 03)	0,092
IMP 11 (MIRRA 04)	0,052
IMP 12 (MIRRA 05)	0,028
IMP 13 (MIRRA 06)	5,290

CT02: Model: Suspect2

FILE	LR
Alert	0,235
IMP 01	0,045
IMP 02	0,035
IMP 03	0,031
IMP 04	0,033
IMP 05	0,033
IMP 06	0,036
IMP 07	0,063
IMP 08 (MIRRA 01)	0,041
IMP 09 (MIRRA 02)	0,032
IMP 10 (MIRRA 03)	0,030
IMP 11 (MIRRA 04)	0,036
IMP 12 (MIRRA 05)	0,040
IMP 13 (MIRRA 06)	0,048

Note.- LR values in the tables range between [0,03 2700] to avoid non-calibrated LRs. This is based on experimental studies using BDRA public file (Order INT/1202/2011, 4th of May, Spanish Ministry of Interior, BOE nº 114, 13th of May, 2011).

Evaluation

On one hand, there have been used the following propositions in CT01:

Proposition 1: Suspect1 is the author of the questioned voice

Proposition 2: Other people is the author of the questioned voice being recorded in similar acoustic conditions

As presented in Table CT01, findings shows a moderately strong support towards the hypothesis which points at suspect1 as the author of the questioned voice versus the hypothesis of the authorship by other people recorded in similar acoustic conditions. Such a result is strengthened when the values obtained by impostors are observed. In spite of it, there is an impostor file recorded in acoustic conditions very close to that of the questioned recording whose LR value is higher than one. In any case, this value is very low and coherent with the misleading evidence rate expected for this system in such working conditions. And, in the other hand, there have been used the following propositions in CT02: Proposition 1: Suspect2 is the author of the questioned voice Proposition 2: Other people is the author of the questioned voice being recorded in similar acoustic conditions With regard to the CT02 comparison, findings slightly support the hypothesis that the questioned voice comes from other people whose voices were recorded in similar acoustic conditions than that of suspect3 versus the hypothesis of authorship by the suspect3. In all these cases, the technology of the speaker recognition system used minimizes misleading evidence consisting of LRs higher than 1 when the suspect is not the author of the questioned speech. Such a property contributes to an increase of misleading evidence in the opposite direction.	
Conclusions 1.- Findings shows a moderately strong support towards the hypothesis which points at Suspect1 as the author of questioned voice versus the hypothesis of the authorship by other people recorded in similar acoustic conditions. 2.- Findings slightly supports the hypothesis that the questioned voice comes from a person different to Suspect2 whose voice was recorded in similar acoustic conditions versus the opposite hypothesis. 3.- Suspect3 could not be compared by technical reasons.	

Interpretation of such values is made according to the following table (H_1 and H_2 are exhaustive and complementary hypotheses):

Verbal scale	LR
Slight support /Limited support H_1	$1 > LR \leq 10$
Moderate support H_1	$10 > LR \leq 100$
Moderately strong support H_1	$100 > LR \leq 1000$
Strong support H_1	$1000 > LR \leq 10000$
Very strong support H_1	$LR > 10000$
Slight support /Limited support H_2	$0.1 \leq LR < 1$
Moderate support H_2	$0.01 \leq LR < 0.1$
Moderately strong support H_2	$0.001 \leq LR < 0.01$
Strong support H_2	$0.0001 \leq LR < 0.001$
Very strong support H_2	$LR < 0.0001$

BIBLIOGRAPHY

AFSP (Association of Forensic Science Providers), Standards for the formulation of evaluative forensic science expert opinion. Science and Justice 49(3), 2009, pp.161-164.

AGAZZI E, Proposta di una nuova caratterizzatione dell'oggettività scientifica. Itinerari 1-2, 1979.

AITKEN CGG, Statistics and Evaluation of Evidence for Forensic Scientists. John Wiley & Sons, 1st edition, Chichester (UK), 1995.

AITKEN CGG, TARONI F, Statistics and Evaluation of Evidence for Forensic Scientists. John Wiley & Sons, 2nd edition, Chichester (UK), 2004.

AITKEN CGG, BERGER C, BUCKLETON J, CHAMPOD C, CURRAN J, DAWID A, EVETT I, GILL P, GONZALEZ RODRIGUEZ J, JACKSON G, KLOOSTERMAN A, LOVELOCK T, LUCY D, MARGOT P, MCKENN L, MEUWLY D, NEUMANN C, DAEID N, NORDGAARD A, PUCH SOLIS R, RASMUSSON B, REDMAYNE M, ROBERTS P, ROBERTSON B, ROUX C, SJERPS M, TARONI F, TJIN A TSOI T, VIGNAUX G, WILLIS S, ZADORA G, Expressing evaluative opinions: A position statement. Science and Justice, Guest editorial, 51, Issue 1, 2011, pp. 1-2.

ALCARAZ VARO E, HUGHES B, Diccionario de Términos Jurídicos. Ariel, 6ª edición, Barcelona (SP), 2001.

ALCON MJ, AMADOR J, CACERES I, GIRON P, NIETO C, PEREZ T, Estimation of the probability of misleading evidence in the case of normal populations with known different variances. Technical Report, Universidad Complutense de Madrid, 2007. Available at http://atvs.ii.uam.es/files/2009_TR_Alcon.pdf (last access on 15th of December, 2014)

ALLEN R, A Note To My Philosophical Friends About Expertise And Legal Systems. Humana.Mente Journal of Philosophical Studies 28, 2015, pp. 71-86.

ALVARADO VELLOSO A, La prueba judicial (Reflexiones críticas sobre la confirmación procesal). Monografía nº 457, Tirant lo blanch, Valencia (SP), 2006.

ALVIRA T et al., Metafísica. EUNSA, 8ª edición. Pamplona (SP), 2001.

ANDERSON T, SCHUM D, TWINING W, Analysis of Evidence. Cambridge University Press, 2nd edition, Cambridge (UK), 2010.

ARANA J, Los sótanos del universo. Biblioteca Nueva, Madrid (SP), 2012.

ARANA J, La conciencia inexplicada. Biblioteca Nueva, Madrid (SP), 2015.

ARISTOTLE, Metaphysics, book IX.

ARISTOTLE, De Anima, book III.

ARTIGAS M, Filosofía de la Ciencia Experimental. EUNSA, 2ª edición ampliada, Pamplona (SP), 1992.

AUCKENTHALER R, CAREY M, LLOYD THOMAS H, Score normalization for text-independent speaker verification systems. Digital Signal Processing, 10(1-3), 2000, pp. 42-54.

BALDING DJ, Weight-of-evidence for Forensic DNA Profiles. John Wiley & Sons, Chichester (UK), 2005.

BEN M, BLOUET R, BIMBOT FA, Monte Carlo method for score normalization in automatic speaker verification using Kullback-Leibler distances. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando (USA), 2002.

BERGER J, BERRY D, Statistical analysis and the illusion of objectivity. American Scientist, 76(2), 1988, pp. 159-165.

BIEDERMANN A, BOZZA S, TARONI F, Probabilistic evidential assessment of gunshot residue particle evidence (Part I): Likelihood ratio calculation and case pre-assessment using Bayesian networks. Forensic Science International, 191(1-3), 2009, pp. 24-35.

BIEDERMANN A, BOZZA S, TARONI F, Probabilistic evidential assessment of gunshot residue particle evidence (Part II): Bayesian parameter estimation for experimental count data. Forensic Science International, 206(1-3), 2010, pp. 103-110.

BIEDERMANN A, The Role of the Subjectivist Position in the Probabilization of Forensic Science. Journal of Forensic Science Medicine, 1 (2), 2015, pp. 140-148.
BIRNBAUM A, On the foundations of statistical inference (with discussion). Journal of the American Statistical Association, 57(298), 1962, pp. 269-306.
BOHAN TL, Strengthening forensic science: a way station on the journey to Justice (President's Editorial). Journal of Forensic Sciences, 55(1), 2010, pp. 5-7.
BRUMMER N, DU PREEZ J, Application independent evaluation of speaker detection. Computer Speech and Language, 20(2-3), 2006, pp. 230-275.
BRUMMER N, GARCIA ROMERO D, Generative Modelling for Unsupervised Score Calibration. Accepted for ICASSP 2014. Available at http://arxiv.org/abs/1311.0707 (accessed on January 12, 2017).
BUCKLETON JS, TRIGGS CM, WALSH SJ (Editors), Forensic DNA Evidence Interpretation. CRC Press, Boca Raton (FL – USA), 2004.
BUNCH SG, Consecutive matching striation criteria: a general critique. Journal of Forensic Sciences, 45, 2000, pp. 955-962.
BUNCH SG, WEVERS G, Application of likelihood ratios for firearm and toolmark analysis. Science and Justice, 53(2), 2013, 223-229.
BUTLER JM, Forensic DNA typing. Elsevier Academic Press, Barlington (MA – USA), 2005.
CALDERON CUADRADO MP, La segunda instancia penal. Thomson-Aranzadi, Navarra (SP), 2005.
CAMPBELL WM, STURIM DE, REYNOLDS DA, Support vector machines using GMM supervectors for speaker verification. IEEE Signal Processing Letters, 13(5), 2006, pp. 308-311.
CHAMPOD C, Reconaissance automatique et analyse statistique des minuties sur les empreintes digitales. PhD thesis, Institut de Police Scientifique et de Criminologie, Université de Lausanne, Imprimerie Evard, Concise (SW), 1996.
CHAMPOD C, TARONI F, Bayesian framework for the evaluation of fibre transfer evidence. Science and Justice, 37(2), 1997, pp. 75-83.
CHAMPOD C, Interpretation of evidence and reporting in the light the 2009 NRC report. Keynote Speech of the Interpretation and Evaluation Session of the V Conference of the European Academy of Sciences held in Glasgow, University of Strathclyde (UK), 2009.
CHAMPOD C, VUILLE J, Scientific evidence in Europe – Admissibility, Appraisal and Equality of Arms. European Committee on Crime Problems (CDCP), Strasbourg (FR), 2010.
CLEMENTE LAZARO FJ, Las piezas de convicción. Bubok Publishing S.L., ISBN: 978-84-92580-45-3, 2000.
COHEN HL, The Probable and the Provable. Clarendon Press, Oxford (UK), 1977.
CORAZON GONZALEZ R, Filosofía del conocimiento. EUNSA, 1ª edición. Pamplona (SP), 2002.
COUNCIL FRAMEWORK DECISION 2009/905/JHA of November 2009 on Accreditation of forensic service providers carrying out laboratory activities. Acts adopted under Title VI of the EU Treaty. Official Journal of the European Union of 9 January 2009, L 322.
CURRAN JM, HICKS TN, BUCKLETON JS, Forensic Interpretation of Glass Evidence. CRC Press, Boca Raton (MA – FL), 2000.
DAVIS S, MERMELSTEIN P, Comparison of parametric representations for monosyllabic Word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 1980, pp. 357-366.
DEHAK N, KENNY P, DEHAK R, DUMOUCHEL P, OUELLET P, Front-end factor analysis for speaker verification. IEEE Transactions on Audio, Speech, and Language Processing, 19(4), 2011, pp. 788-798.
DE FINETTI B, Fondamenti logici del ragionamento probabilistico. Bolletino Della Regia Unione Matematica Italiana, 9, 1930.
DE TORRE JM, Christian Philosophy. SINAG-TALA, Manila (PH), 1980.

DIACONIS P, FREEDMAN D, The persistence of cognitive illusions. Journal of Behavioral and Brain Science, 4, 1981, pp. 333-334.
DIAZ CABIALE JA, LOPEZ CASTILLO M, La conversión de la prueba pericial en documental, artículo 788.2 II LECr. Jueces para la democracia 46, 2003, p. 67.
DROR IE, CHARLTON D, Why experts make errors. Jounal of Forensic Identification, 56, 2006, pp. 600-610.
EINSTEIN A, L'évolution des idées du physique. Payot, Paris (FR), 1978.
ELLMAN IM, KAYE D, Probabilities and Proof: Can HLA and Blood Group Testing Prove Paternity? New York University Law School, 54, Rev. 1131, 1979, pp. 1131-1162.
ENFSI DRUGS WORKING GROUP. Wiesbaden: Proficiency Test organized by Dr. Wolf-Rainer Bork (Bundeskriminalamt), 2008-2009.
ENFSI GUIDELINE FOR EVALUATIVE REPORTING IN FORENSIC SCIENCE. March, 2015.
EVETT IW, A quantitative theory for interpreting transfer evidence in criminal cases. Applied Statistics, 33, 1984, pp. 25-32.
EVETT IW, WEIR BS, Interpreting DNA Evidence. Sinauer Associates Incorporated, Sunderland (MA - USA), 1988.
EVETT IW, Expert evidence and forensic misconceptions of the nature of exact science. Science and Justice, 36(2), 1996, pp. 118-122.
EVETT IW, BUCKLETON JS, Statistical analysis of STR data. Advances in Forensic Haemogenetics, Springer-Verlag, Heildeberg 6, 1996b, pp. 79-86.
EVETT IW, LAMBERT JA, BUCKLETON JS, A Bayesian approach to interpreting footwear marks in forensic casework. Science and Justice, 38, 1998, pp. 241-247.
EVETT IW, Evaluation and professionalism. Science and Justice, 49 (3), 2009, pp. 159-160.
FIENBERG SE, KAYE DH, Legal and statistical aspects of some mysterious clusters. Journal of the Royal Statistical Society, Series A, 154, 1991, pp. 265-270.
FORMENT E, Metafísica. Colección Albatros. Palabra, Madrid (SP), 2009.
FORSTAT: http://www.ies.krakow.pl/conferences/forstat2016 (accessed on January 12, 2017).
FRIEDMAN RD, Assessing Evidence. Michigan Law Rev, 94, 1996, pp.1836-37.
FURUI S, Cepstral analysis technique for automatic speaker verification. IEEE Transactions on Acoustics, Speech, and Signal Processing, 29(2), 1981, pp. 254-272.
GARCIA ROMERO D, ESPY WILSON CY, Analysis of i-vector length normalization in speaker recognition systems. Proceedings of the 12th INTERSPEECH Conference, 2011, pp. 249-252.
GASCON ABELLAN M, Los hechos en el derecho. Marcial Pons, 3ª edición, Madrid (SP), 2010.
GASCON ABELLAN M, LUCENA MOLINA JJ, GONZALEZ RODRIGUEZ J, Razones científico-jurídicas para valorar la prueba científica: una argumentación multidisciplinar. Diario La Ley, nº 7481, Sección Doctrina, 2010a.
GASCON ABELLAN M, LUCENA MOLINA JJ, Pruebas científicas: necesidad de un cambio de paradigma. Revista Jueces para la Democracia, nº 69, 2010b, pp. 95-106.
GONZALEZ RODRIGUEZ J, DRYGAJLO A, RAMOS CASTRO D, GARCIA GOMAR M, ORTEGA GARCIA J, Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. Computer Speech and Language, 20(2-3), 2006, pp. 331-355.
GONZALEZ RODRIGUEZ J, ROSE P, RAMOS CASTRO D, TOLEDANO DT, ORTEGA GARCIA J, Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. IEEE Transactions on Audio, Speech and Language Processing, 15(7), 2007, pp. 2104-2115.
GONZALEZ RODRIGUEZ J, Evaluating Speaker Recognition systems: An overview of the NIST Speaker Recognition Evaluations (1996-2014). Loquens, 1(1), e007, 2014. Available at:

http://loquens.revistas.csic.es/index.php/loquens/article/download/9/21 (accessed on January 12, 2017)

GOOD PI, HARDIN JW, Common errors in statistics (and how to avoid them). J. Wiley & Sons, Hoboken (NJ-USA), 2003.

GOVERNMENT OF THE UNITED KINGDOM, The Law Commission Consultation Paper No 190, The admissibility of expert evidence in criminal proceedings in England and Wales - A new approach to the determination of evidentiary reliability. Available at: http://www.lawcom.gov.uk/wp-content/uploads/2015/03/cp190_Expert_Evidence_Consultation.pdf (accessed on January 12, 2017).

GUASTINI R, L'interpretazione revisitata. Distinguiendo. Studi di teoria e metateoría del diritto, Torino (IT), 1996.

GUERRERO ZAPLANA J, Valoración de la prueba pericial en el juicio oral. Presentation in the II Symposium on Forensic Speaker Recognition organized by the Universidad Politécnica de Cartagena (SP), 1999.

HACKING I, The Logic of Statistical Inference. Cambridge University Press, Cambridge (UK), 1965.

HERMANSKY H, MORGAN N, RASTA processing of speech. IEEE Transactions on Speech and Audio Processing, 2(4), 1994, pp. 578-589.

INNOCENCE PROJECT: www.innocenceproject.org (accessed on January 12, 2017).

INTERPOL EUROPEAN EXPERT GROUP ON FINGERPRINT IDENTIFICATION, Method for Fingerprint Identification, Part 2: Detailing the method using common terminology and through the definition and application of shared principles. Available at: http://www.interpol.int/Public/Forensic/Fingerprints/WorkingParties/IEEGFI2/IEEGF2.pdf (last access on January 17, 2011)

JAYNES ET, Probability theory, the logic of science. Cambridge University Press, 2003.

JEFFREY RC, The Logic of Decision. University of Chicago Press, 2nd edition, 1983.

JEFFREYS H, Scientific inference. Cambridge University Press, 1931.

KAYE DH, Probability, individualization, and uniqueness in forensic science: listening to the Academies. Social Science Research Network (SSRN), 2009. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1261970 (accessed on January 12, 2017).

KENNEDY D, Forensic science: Oxymoron? (editorial). Science 302, 2003, p. 1625.

KENNEDY RB, PRESSMAN IS, CHEN S, PETERSEN PH, PRESSMAN AE, Statistical analysis of barefoot impressions. Journal of Forensic Sciences, 48, 2003, pp. 53-63.

KENNY P, Joint Factor Analysis of speaker and session variability: Theory and algorithms. Technical Report No. CRIM-06/08-13, Montreal (Canada), 2005.

KENNY P, OUELLET P, DEHAK N, GUPTA V, DUMOUCHEL P, A Study of Inter-Speaker Variability in Speaker Verification. IEEE Transactions on Audio, Speech Lang. Process., 16(5), 2008, pp. 980-988.

KINGSTON CR, KIRK PL, The use of statistics in criminalistics. Journal of criminal law, criminology and police science, 5, 1964, pp. 514-521.

LAD F, Operational Subjective Statistical Methods: a Mathematical, Philosophical, and Historical Introduction. J. Wiley & Sons, New York (USA), 1996.

LAURITZEN SL, SPIEGELHALTER DJ, Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society, Series B 50(2), 1988, pp. 157-224.

LINDLEY DV, Introduction to Probability and Statistics from a Bayesian Viewpoint. Cambridge University Press, Cambridge (UK), 1965.

LINDLEY DV, Understanding Uncertainty. J. Wiley & Sons, 2nd edition, Hoboken (NJ-USA), 2014.

LUCENA MOLINA JJ, PARDO IRANZO V, ESCOLA GARCIA MA, Elementos para el debate sobre la valoración de la prueba científica en España: hacia un estándar acreditable bajo la norma 17.025 sobre conclusiones de informes periciales. Revista Internacional de Estudios sobre Derecho Procesal y Arbitraje, Nº 2, 2011. Available at: www.riedpa.com (accessed on January 12, 2017)).

LUCENA MOLINA JJ, PARDO IRANZO V, GONZALEZ RODRIGUEZ J, Weakening Forensic Science in Spain: From Expert Evidence to Documentary Evidence. Journal of Forensic Sciences, 57(4), July 2012, pp. 952-963.

LUCENA MOLINA JJ, GASCON ABELLAN M, PARDO IRANZO V, Technical support for a judge when assessing a priori odds. Law, Probability and Risk, 14(2), June 2015, pp. 147-168.

LUCENA MOLINA JJ, RAMOS CASTRO D, GONZALEZ RODRIGUEZ J, Performance of likelihood ratios considering bounds on the probability of observing misleading evidence. Law, Probability and Risk, 14(3), September 2015, pp. 175-192.

LUCENA MOLINA JJ, Epistemology applied to conclusions of expert reports. Forensic Science International, Special Issue EAFS 2015, 264, July 2016, pp. 122-131.

LLANO A, Gnoseología. EUNSA, 1ª edición, Pamplona (SP), 2007.

McDERMOTT SD, WILLIS SM, A survey of the evidential value of paint transfer evidence. Journal of Forensic Sciences, 42, 1997, pp. 1012-1018.

McDERMOTT SD, WILLIS SM, McCULLOUGH JP, The evidential value of paint. Part II: A Bayesian approach. Journal of Forensic Sciences, 44, 1999, pp. 263-269.

MEANS OF PROOF: http://ec.europa.eu/civiljustice/evidence/evidence_spa_es.htm (accessed January 12, 2017)

MEUWLY D, Reconaissance de Locuteurs en Sciences Forensiques: L'apport d'une Approache Automatique. PhD. Thesis, IPSC-Universite de Lausanne, 2001.

MILLAN PUELLES A, Léxico Filosófico. RIALP, 2ª edición, Madrid (SP), 2002.

MOROS ER, UMBERS RJ, ¿Qué es el conocimiento? La epistemología en los Estados Unidos hoy. Anuario Filosófico, XXXVI/3, Pamplona (SP), 2003, pp. 633-671.

NATIONAL ACADEMY OF SCIENCES, On the Theory and Practice of Voice Identification. National Technical Information Service of the US Department of Commerce, Washington DC, 1979.

NATIONAL RESEARCH COUNCIL, The National Academies, Strengthening Forensic Science in the United States: a path forward. National Academies Press, Washington DC (USA), 2009.

NEUMANN C, CHAMPOD C, PUCH SOLIS R, EGLI N, ANTHONIOZ D, BROMAGE GRIFFITHS A, Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. Journal of Forensic Sciences, 52(1), 2007, pp. 54-64.

NORGAARD A, ANSELL R, JAEGER L, DROTZ W, Ordinal scales of conclusion for the value of evidence. Presentation in the Interpretation and Evaluation Session of the V Conference of the European Academy of Sciences held in Glasgow, University of Strathclyde (UK), 2009.

O'HAGAN A, BUCK CE, DANESHKHAH A, EISER JR, GARTHWAITE PH, JENKINSON DJ, OAKLEY JE, RAKOW T, Uncertain judgements: eliciting expert's probabilities. Wiley, Hoboken (NJ-USA), 2006.

PARDO IRANZO V, La prueba documental en el proceso penal. Tirant lo blanch, Valencia (SP), 2008.

PARKER JB, HOLFORD A, Optimum test statistics with particular reference to a forensic problem. Applied statistics, 17, 1968, pp. 237-251.

PEARL J, Reverend Bayes on inference engines: A model of self-activated memory for evidential reasoning. Proceedings of the Cognitive Science Society, Ablex, Greenwich (UK), 1982, pp. 133-136.

PELECANOS J, SRIDHARAN S, Feature warping for robust speaker verification. Proceedings of 'A speaker odyssey: The speaker recognition workshops', 2001, pp. 213-218.

PIERRINI G, DOYLE S, CHAMPOD C, TARONI F, WAKELIN D, LOCK C, Evaluation of preliminary isotopic analysis (^{13}C and ^{15}N) of explosives. A likelihood ratio approach to assess the links between Semtex samples. Forensic Science International, 167, 2007, pp. 43-48.

PLUMPE MD, QUATIERI TF, REYNOLDS DA, Modelling of the glottal flow derivative waveform with application to speaker identification. IEEE Transactions on Speech and Audio Processing, 7(5), 1999, pp. 569-586.

PRINCE SJD, ELDER JH, Probabilistic linear discriminant analysis for inferences about identity. IEEE 11th International Conference on Computer Vision, 2007, pp. 1-8.

RAMOS CASTRO D, Forensic evaluation of the evidence using automatic speaker recognition systems. PhD thesis, Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid (SP), 2007. Available online at: http://atvs.ii.uam.es (accessed on January 12, 2017).

RAMOS CASTRO D, GONZALEZ RODRIGUEZ J, GONZALEZ DOMINGUEZ J, LUCENA MOLINA JJ, Addressing database mismatch in forensic speaker recognition with Ahumada III: a public real-case database in Spanish. Proceedings of Interspeech, 2008, pp. 1493-1496.

RAMOS CASTRO D, GONZALEZ RODRIGUEZ J, Reliable Support: Measuring Calibration of Likelihood Ratios. Forensic Science International, 230, 2013a, pp. 156-159.

RAMOS CASTRO D, GONZALEZ RODRIGUEZ J, ZADORA G, AITKEN C, Information-theoretical assessment of the performance of likelihood ratio computation methods. Forensic Science International, 58(6), 2013b, pp. 1503-1518.

RAMSEY FP, Truth and probability. The Foundations of Mathematics and Other Logical Essays, ed. Braithwaite RB, 1931.

REYNOLDS DA, QUATIERI TF, DUNN RB, Speaker verification using adapted Gaussian mixture models. Digital Signal Processing, 10(1-3), 2000, pp. 19-41.

ROBBINS H, "Statistical methods related to the law of the iterated logarithm". Annals of Mathematical Statistics, 41, 1397-1409, 1970, pp. 1397-1409.

ROBERTS P, AITKEN C, The Logic of Forensic Proof: Inferential Reasoning in Criminal Evidence and Forensic Science. Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, prepared under auspices of the Royal Statistical Society's Working Group on Statistics and the Law, United Kingdom, 2012.

ROBERTSON B, VIGNAUX GA, Probability – the logic of the law. Oxford Journal of Legal Studies, 13, 1993, pp.457-478.

ROBERTSON B, VIGNAUX GA, Interpreting Evidence – Evaluating Forensic Science in the Courtroom. John Wiley & Sons, Chichester (UK), 1995.

ROYALL R, Statistical Evidence, A likelihood paradigm. Chapman & Hall/CRC, London – New York, 1997.

ROYALL R, On the probability of observing misleading statistical evidence. Journal of the American Statistical Association, 95, 2000, pp. 760-780.

SAKS MJ, KOEHLER JJ, What DNA 'fingerprints' can teach the law about the rest of forensic science. Cardozo Law Rev 13, 1991, pp. 361-372.

SAKS MJ, KOEHLER JJ, The coming paradigm shift in forensic identification. Science, 309(5736), 2005, pp. 892-895.

SAKS MJ, KOEHLER JJ, The individualization fallacy in forensic science evidence. Vanderbilt Law Rev, 61(1), 2008, pp. 199-219.

SANGUINETI JJ, Lógica. EUNSA, 7ª edición, Pamplona (SP), 2007.

SANGUINETI JJ, Neurociencia y filosofía del hombre. PALABRA, Pamplona (SP), 2014.
SAVAGE LJ, The Foundations of Statistics. John Wiley and Sons, 1ª ed., New York (USA), 1954.
SCHÖLKOPF B, SMOLA AF, Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge (MA - USA), 2002.
SOLOMONOFF A, CAMPBELL WM, BOARDMAN I, Advances in channel compensation for SVM speaker recognition. Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05), 1, 2005, pp. 629-632.
SPINNEY L, "Science in court: The fine print". Nature 464, 2010, pp. 344-346.
STACEY RB, Report of the erroneous fingerprint individualization in the Madrid train bombing case. Journal of Forensic Identification, 54(6), 2004, pp. 706-715.
STONEY DA, What made us ever think we could individualize using statistics? Journal of Forensic Science Society, 31(2), 1991, pp. 197-199.
TAPIAS RB, Sistemas forenses de reconocimiento automático de locutor: determinación y análisis de sus valores más críticos. Proyecto Fin de Carrera, ETSIT (Escuela Técnica Superior de Ingeniería de Telecomunicación) de la UPM (Universidad Politécnica de Madrid), 2005.
TARONI F, AITKEN CGG, GARBOLINO P, De Finetti's subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary for forensic scientists. Science and Justice, 41(3), 2001, pp. 145-150.
TARONI F, BUCKLETON JS, Likelihood ratio as a relevant and logical approach to assess the value of shoeprint evidence. Information Bulletin for Shoeprint/Toolmark Examiners, 8(2), 2002, pp. 15-25.
TARONI F, AITKEN C, GARBOLINO P, BIEDERMANN A, Bayesian Networks and Probabilistic Inference in Forensic Science. 1st edition, J. Wiley & Sons, Statistics in Practice, Chichester (UK), 2006.
TARONI F, BOZZA S, BIEDERMANN A, GARBOLINO P, AITKEN C, Data Analysis in Forensic Science. J. Wiley & Sons, Statistics in Practice, Chichester (UK), 2010.
TARONI F, BIEDERMANN A, VUILLE J, MORLING N, Whose DNA is this? How relevant a question? (a note for forensic scientists). Forensic Science International: Genetics, 7(4), 2013, pp. 467-470.
TARONI F, BIEDERMANN A, BOZZA S, GARBOLINO P, AITKEN C, Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science. 2nd edition, J. Wiley & Sons, Statistics in Practice, Chichester (UK), 2014.
TARONI F, BOZZA S, BIEDERMANN A, AITKEN C, Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. Law, Probability and Risk, 15(1), March 2016, pp. 1-16.
TARUFFO M, La prueba de los hechos. Trotta, Madrid (SP), 2004.
TARUFFO M, La prueba. Marcial Pons, Madrid (SP), 2008.
THOMAS AQUINAS St., The Summa of St. Thomas Aquinas (Summa theologiae), Benzinger, New York, 1948.
THOMAS AQUINAS St., On the Truth of the Catholic Faith (Summa contra gentiles). Imge Books, Doubleday, New York, 1955.
THOMPSON WC, SCHUMMAN EL, Interpretation of statistical evidence in criminal trials, the prosecutor's fallacy and the defense attorney's fallacy. Law and Human Behavior, 11, 1987, pp. 167-187.
TRIBE LH, Trial by Mathematics. Harvard Law Review, 84, 1971.
VAN LEEUWEN D, BRUMMER N, The distribution of calibrated likelihood-ratios in speaker recognition. Interspeech 2013. Available at http://arxiv.org/abs/1304.1199 (accessed on January 12, 2017).
VELARDE J, Incertidumbre y grados de creencia. Teorema. Vol. XXIV/2, 2005.

VERGEER P, BOLCK A, PESCHIER LJC, BERGER CEH, HENDRIKSE JN, Likelihood ratio methods for forensic comparison of evaporated gasoline residues. Science and Justice, 54(6), 2014, pp. 401-411.

VERNAUX R, Crítica del conocimiento. Herder, Barcelona (SP), 1971.

WEIR BS, Genetic Data Analysis II. Sinauer Associates Incorporated, Sunderland (MA – USA), 1996.

WILLIS S, Development and implementation of an ENFSI standard for reporting evaluative forensic evidence. Monopoly Project (2012-2014), ENFSI, Forensic Science Laboratory, Ireland.

ZADORA G, MARTYNA A, RAMOS D, AITKEN C, Statistical Analysis in Forensic Science – Evidential Value of Multivariate Physicochemical Data. John Wiley & Sons, 1st edition, Chichester (UK), 2014.

ZIEBA PALUS J, ZADORA G, MILCZAREK JM, Differentiation and evaluation of evidence value of styrene acrylic urethane topcoat car paints analysed by pyrolisis-gas chromatography. Journal of Chromatography, A 1179, 2008, pp. 47-58.